

# What can we do if there are no good targets

*Mathematical and Computational Biology in Drug Discovery  
(MCBDD) Module II*

*Dr. Jitao David Zhang  
March-April 2022*

# Do numbers of SNPs and conserved genome fit?

How can it be that individual genomes show only limited numbers of differences, while much of the genome is *junk*, i.e. not constrained by evolution, the loss or gain of which does not seriously affect fitness of the host organism?

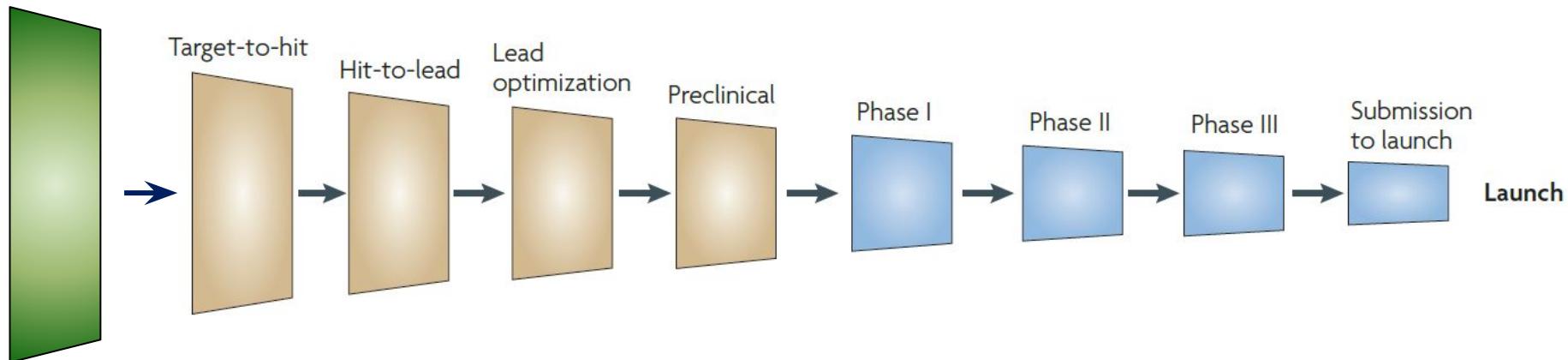
## Facts:

- On average, about 0.6% of the genome of an individual differs from the reference human genome (~5 million sites, affecting 20 million bases).
- Aggregating all known SNPs, we have detected 0.3 billion variants in sequenced samples.
- GWAS usually probes a subset of SNPs in linkage disequilibrium to identify causal variants. Therefore the intra-individual differences may be underestimated.

Numbers for <i>Homo sapiens</i>	Estimate	Source
History	~6.5E5 y	<a href="#">Timeline of the human condition</a>
Genome mutation rate	~1E-9/site/year	<a href="#">Lynch 2010, PNAS</a>
Genome size	~6E9 sites	<a href="#">Human genome hg38</a>
Estimated total mutations in the genome	~6.5E5 * 1E-9 * 6E9 * ~ 4E6, versus 5E6	<a href="#">Leypold 2021, Auton 2015</a>
<del>Effective population size</del>	<del>~1E4 individuals</del>	<a href="#">Charlesworth 2009, Park 2011</a>

# The linear view of drug discovery builds on target-based approaches

Target identification & assessment

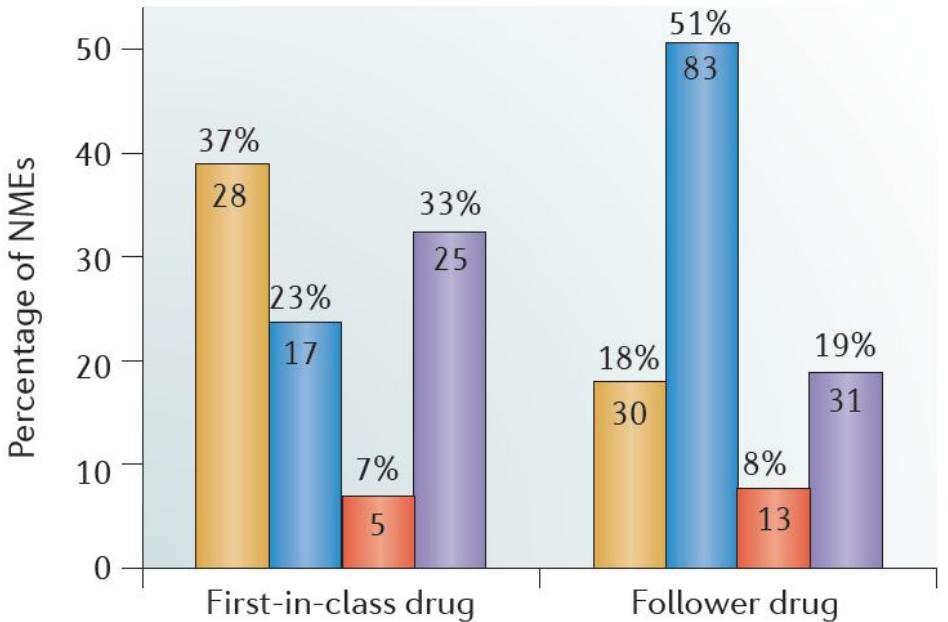
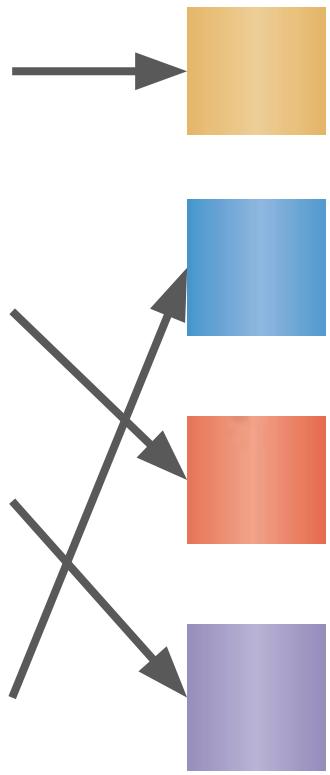


# Five strategies when no good target is found

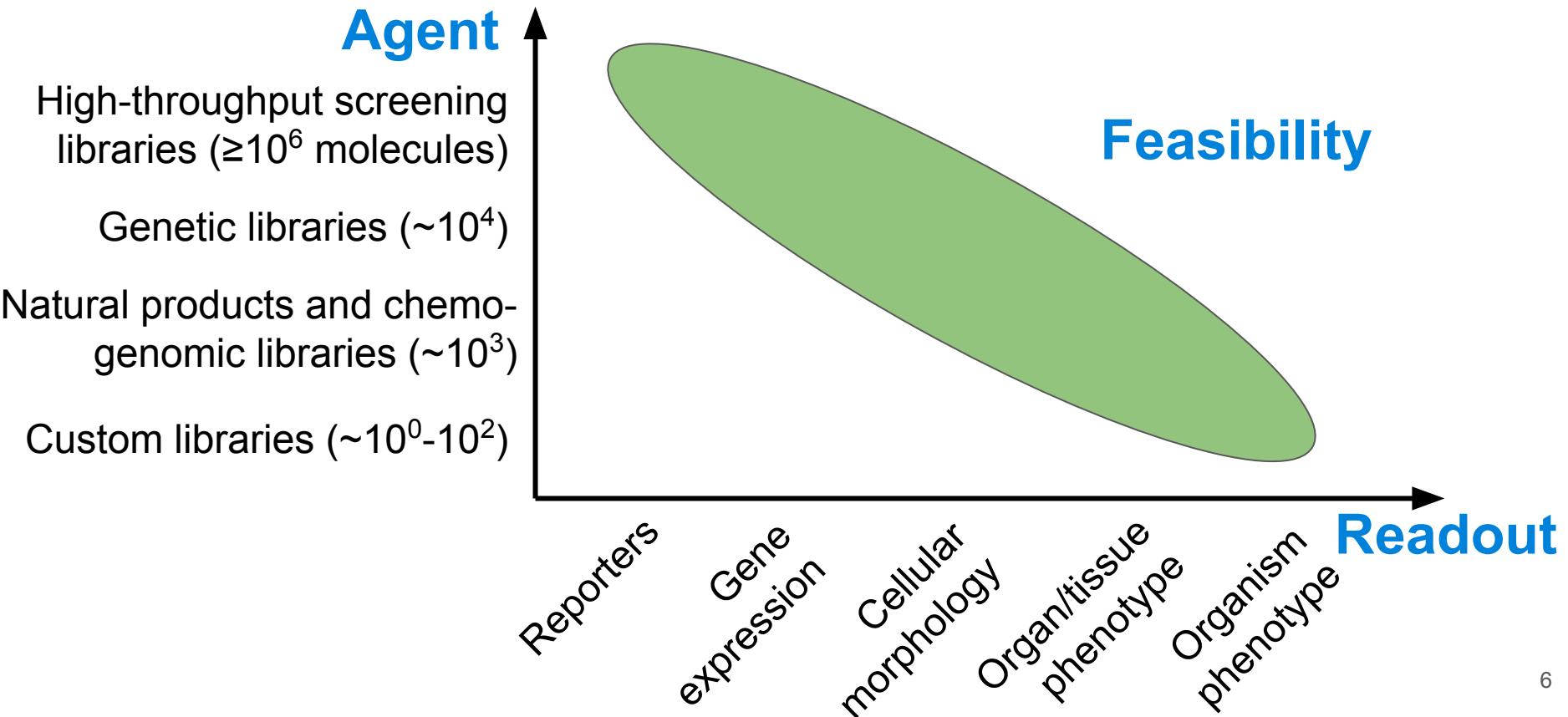
1. Phenotypic drug discovery
2. Natural products
3. Biologics
4. Interaction-based (multispecific) drug discovery
5. Drug repurposing or combination studies

# Connect the lines!

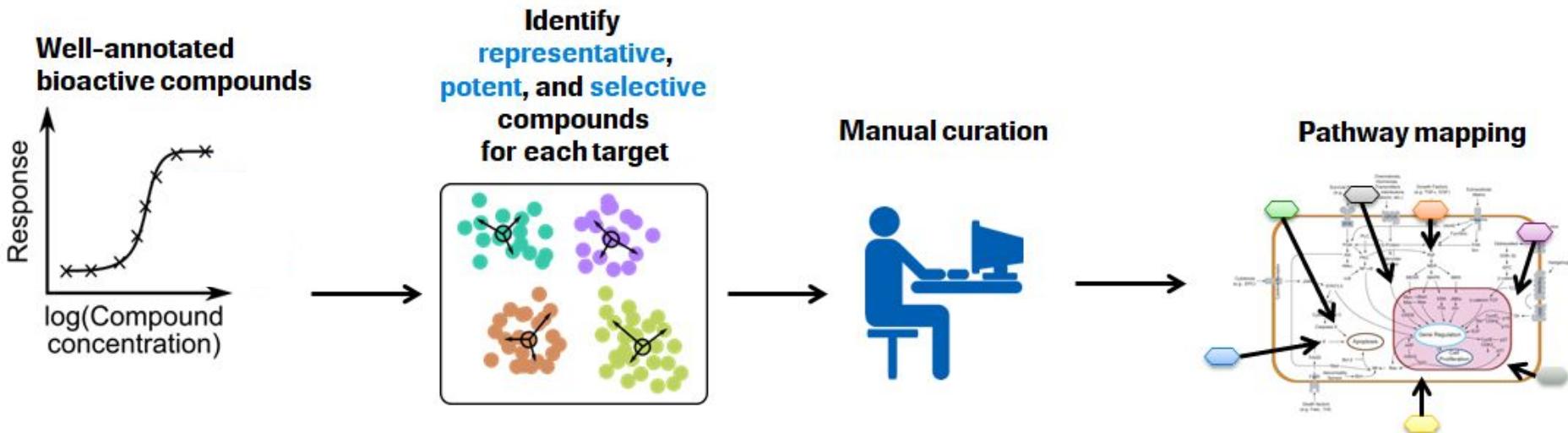
- Phenotypic screening
- Modified natural products
- Biologics
- Target-based screening



# Phenotypic screenings by agent and readout

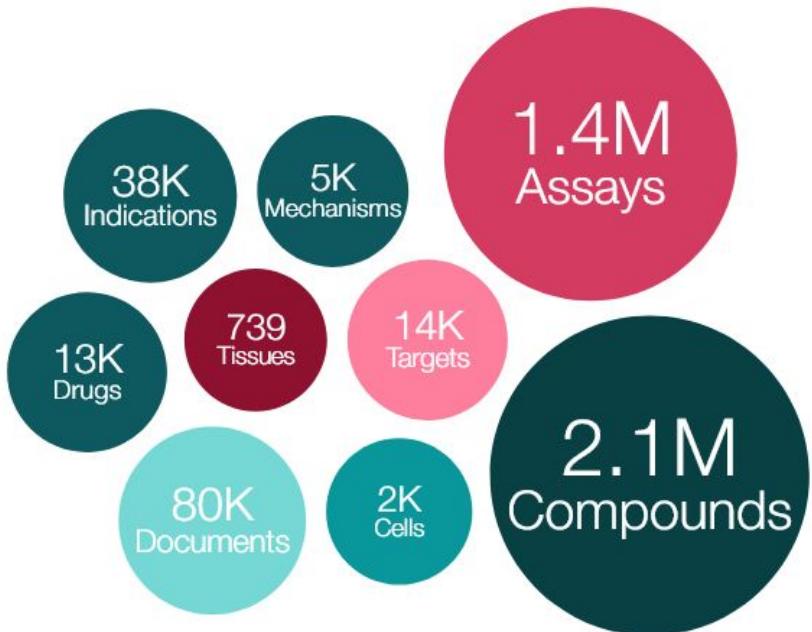


# The Small-molecule PAthway Research Kit (SPARK)



# The ChEMBL database

- An example of query: [aspirin](#).
- Systematic and programmatic accession via [ChEMBL API](#) ([source code](#)).
- We can use **dose-response data** to annotate the *triplets* of compound, assay activity, and targets.

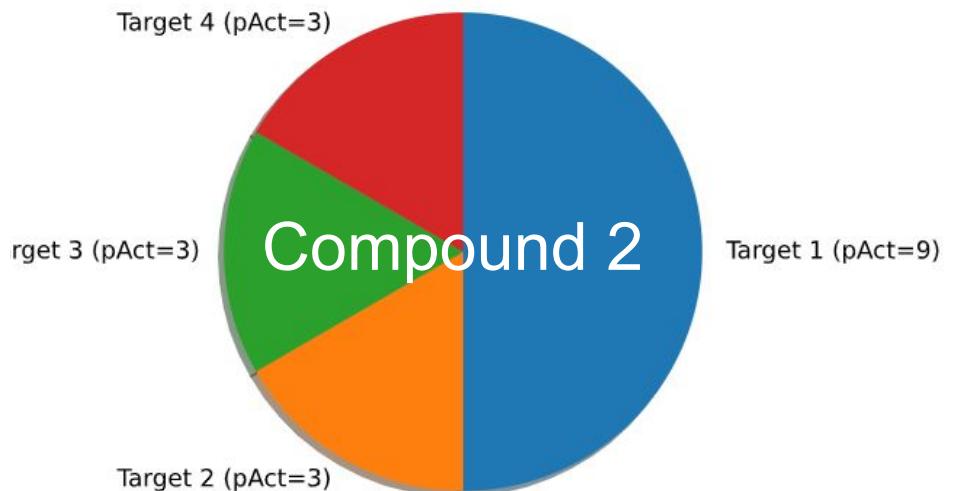
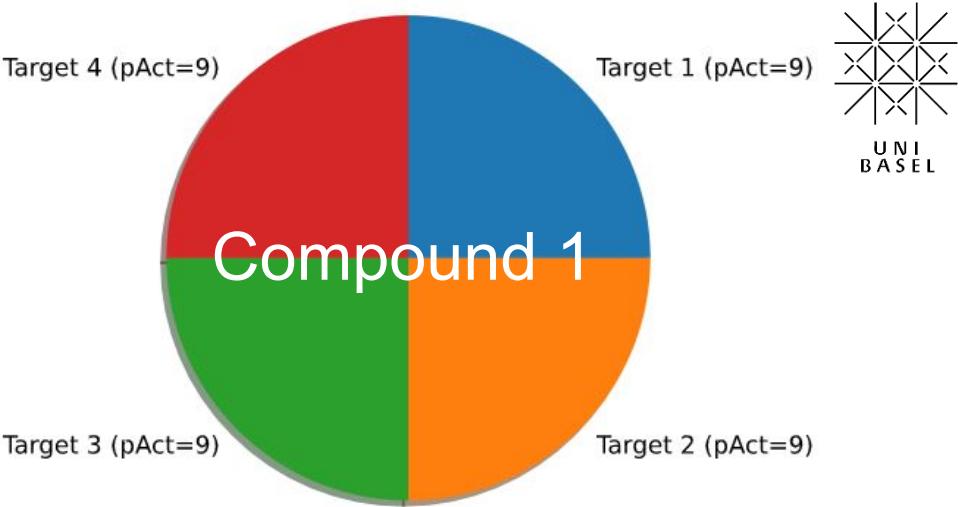
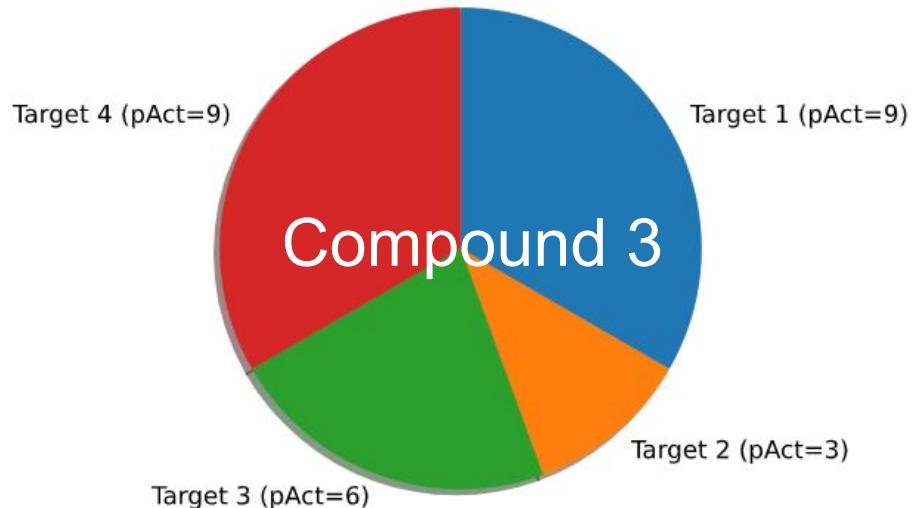


March 2021

# Discussion

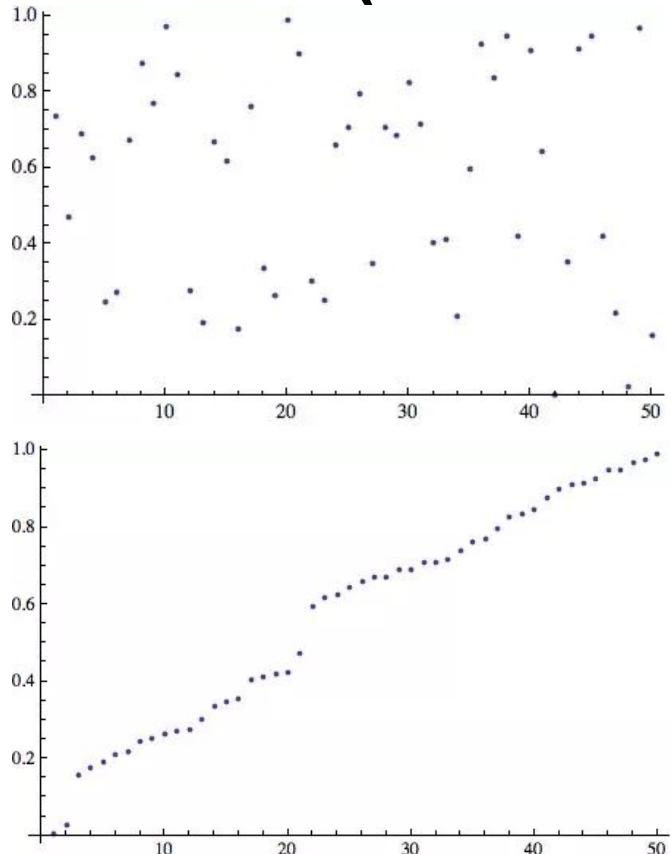
1. Why do we care selecting *representative*, *potent*, and *selective* compounds for each target?
2. How to define following terms mathematically ...
  - a. Representativity?
  - b. Potency?
  - c. Selectivity?

# A toy example about how to quantify a compound's potency and selectivity



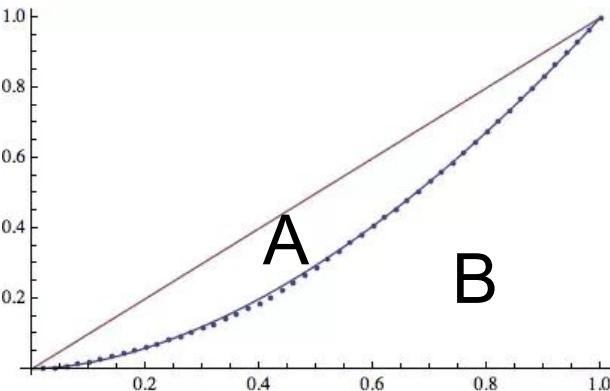
# The Gini Index (a.k.a. Gini Coefficient)

A random vector of 50 values



Sorted from low to high

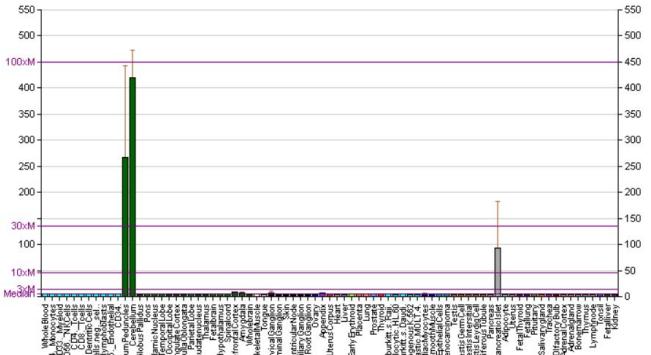
The Gini Index is calculated based on the cumulative distribution



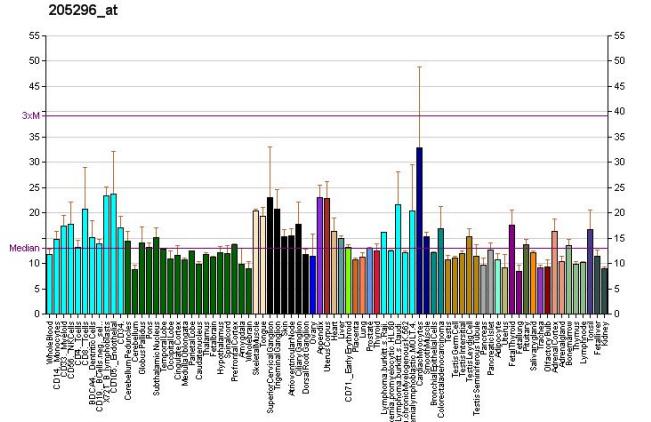
$$G = A/(A+B)$$

# The Gini Index quantifies inequality/ selectivity

*NEUROD1*

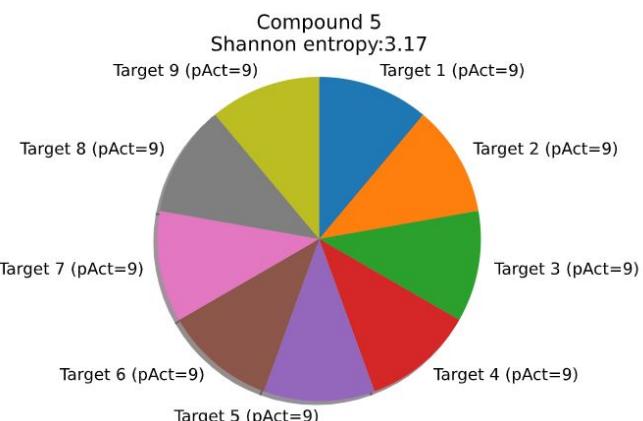
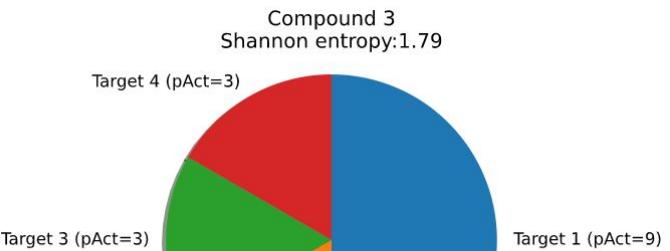
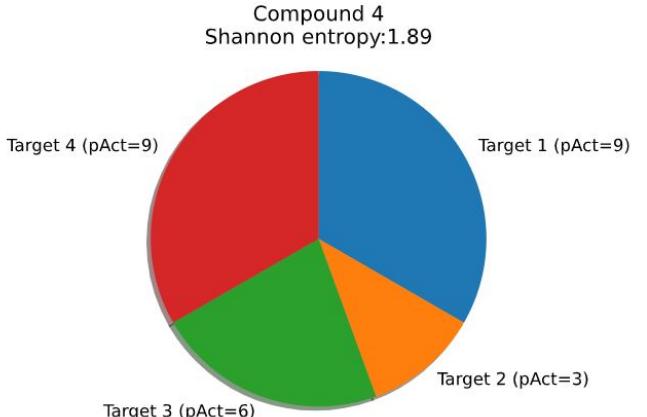
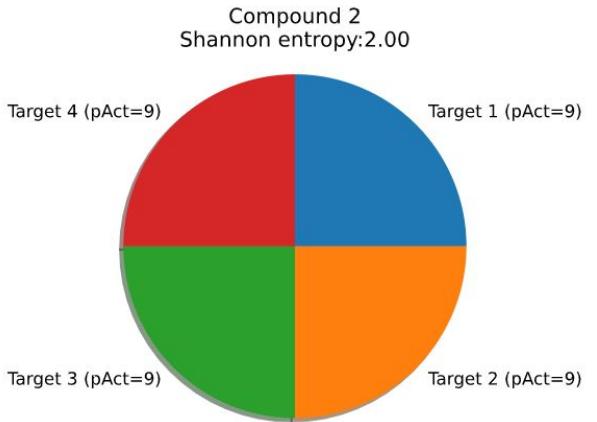
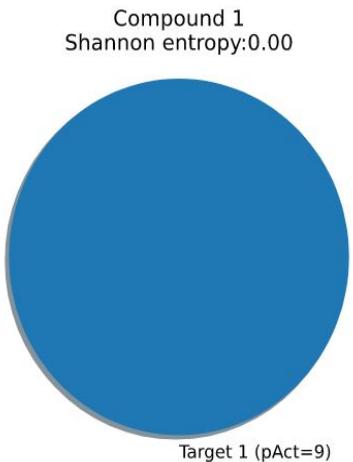


*RBL1*

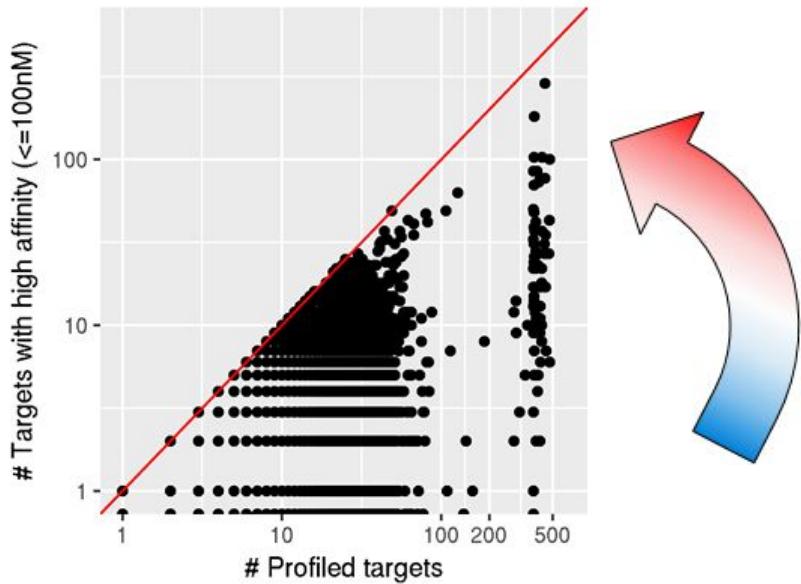


The Gini Index of expression of *NEUROD1* across tissues is near 1, whereas that of *RBL1* is near 0.

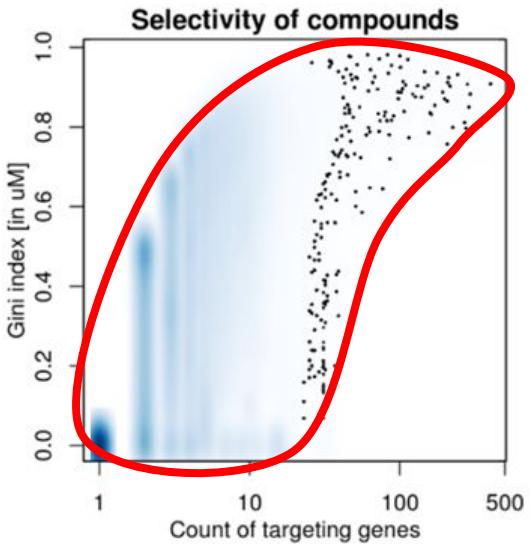
# An alternative metric: Shannon's Entropy



# Count of targets and selectivity of ChEMBL molecules

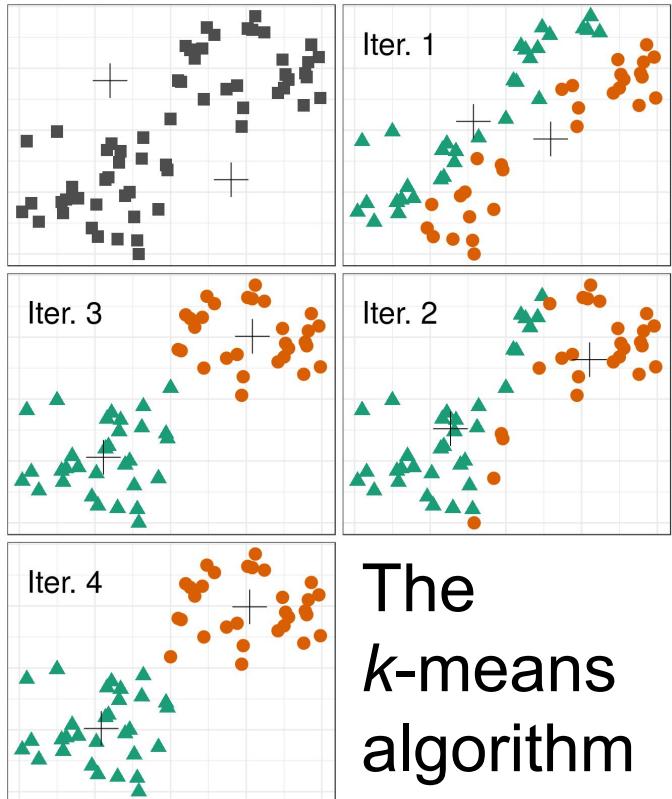


With some exceptions, most compounds are profiled against  $< 100$  targets. We distinguish between specific and pleiotropic compounds.



The **shark-fin shape** curve suggests that frequently profiled compounds tend to be more selective (and *vice versa*).

# Unsupervised clustering

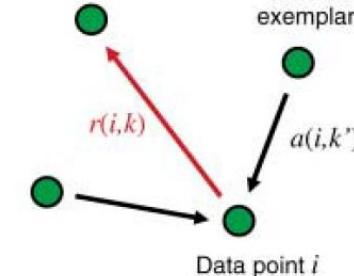


The  
k-means  
algorithm

B

Sending responsibilities

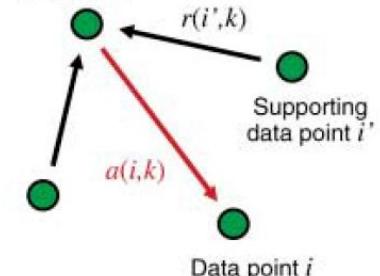
Candidate exemplar  $k$



C

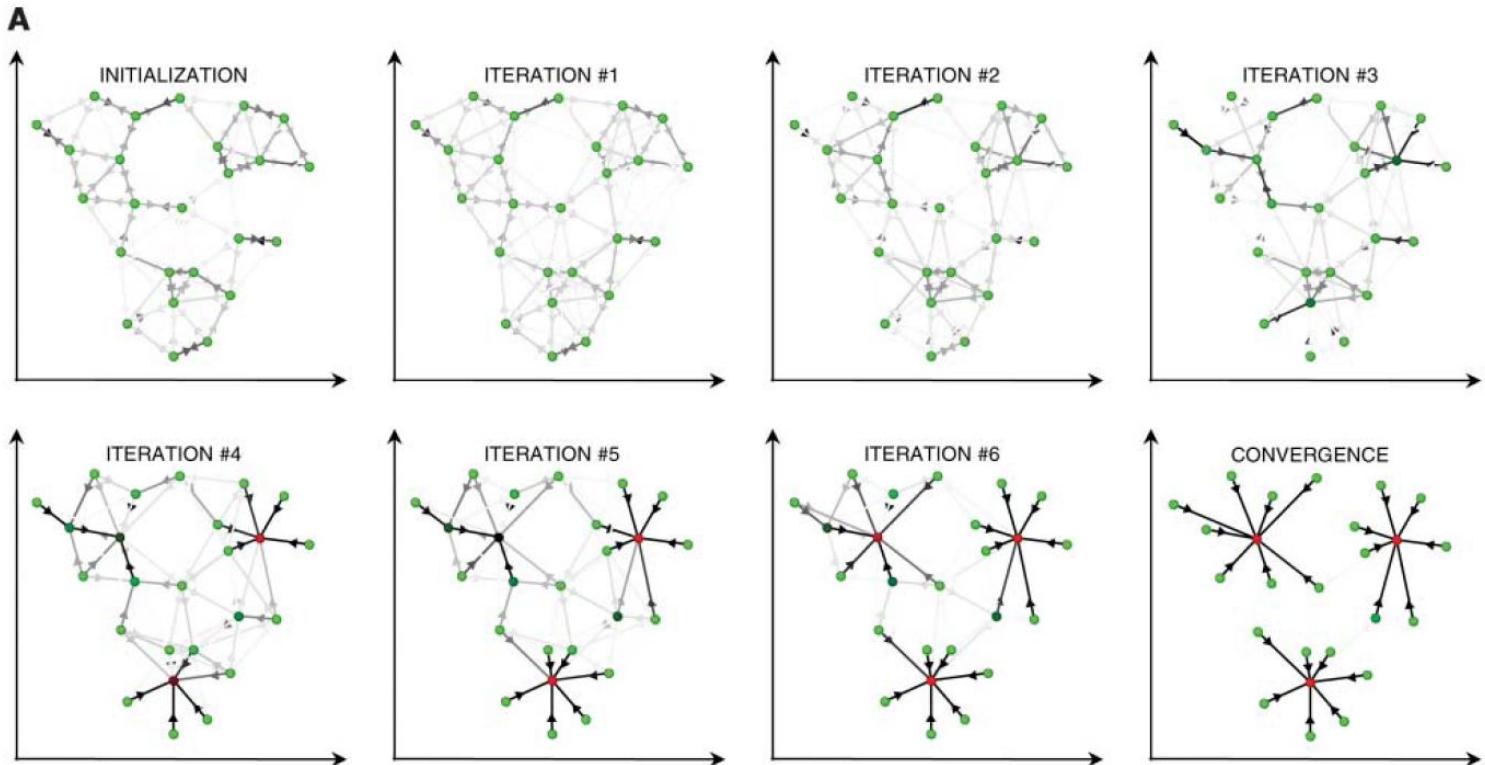
Sending availabilities

Candidate exemplar  $k$



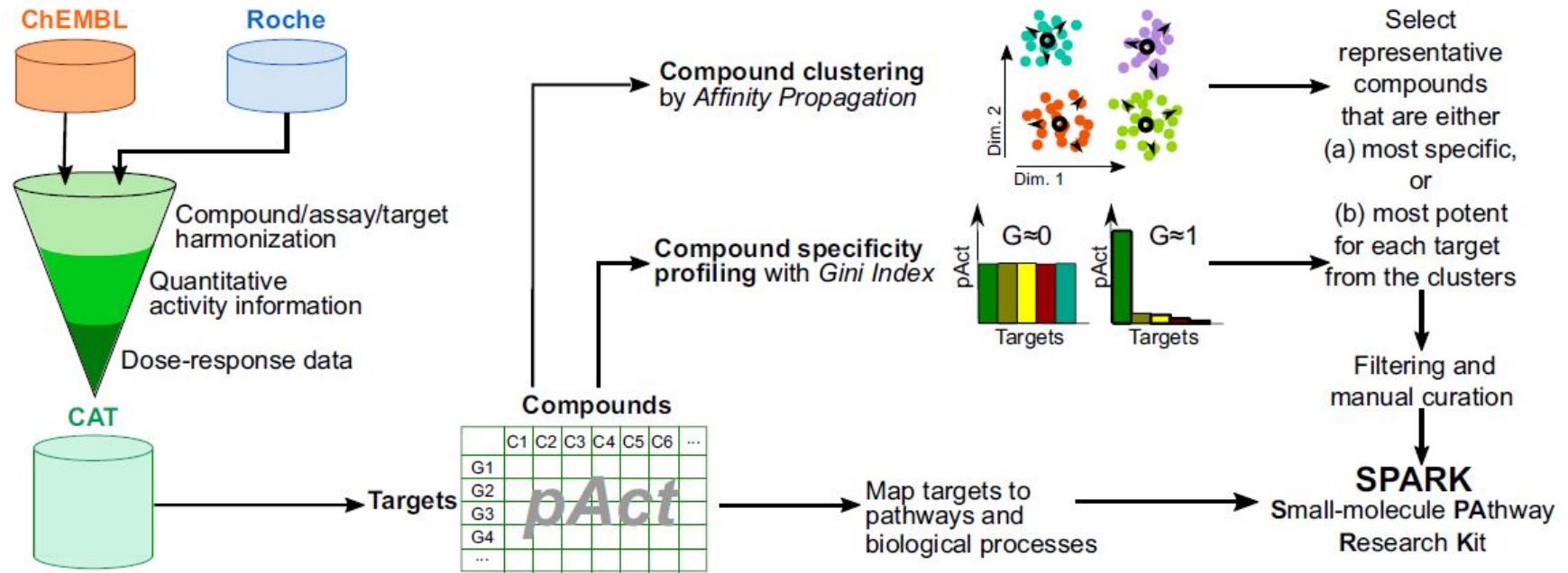
Affinity Propagation updates  
**responsibilities** and  
**availabilities** iteratively

# Affinity Propagation in action



A movie of iterations

# Construction of SPARK in detail



## Harmonization

... of public and  
Roche internal data

## Machine learning

... to select  
compounds

## Pathways

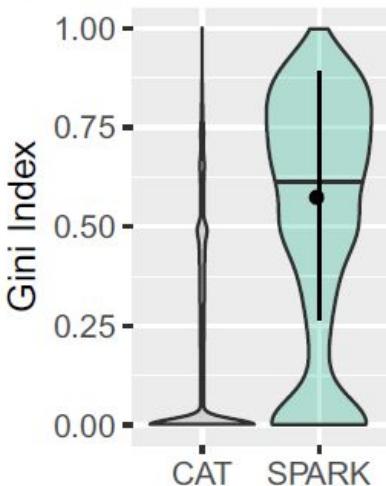
... mapped to  
compounds

## Curation

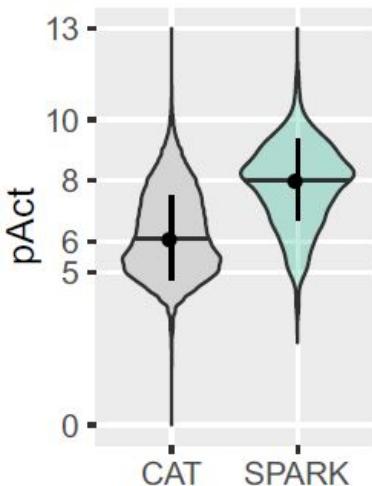
... to enrich quality  
compounds

# SPARK covers the chemical space evenly with representative, potent, and specific compounds

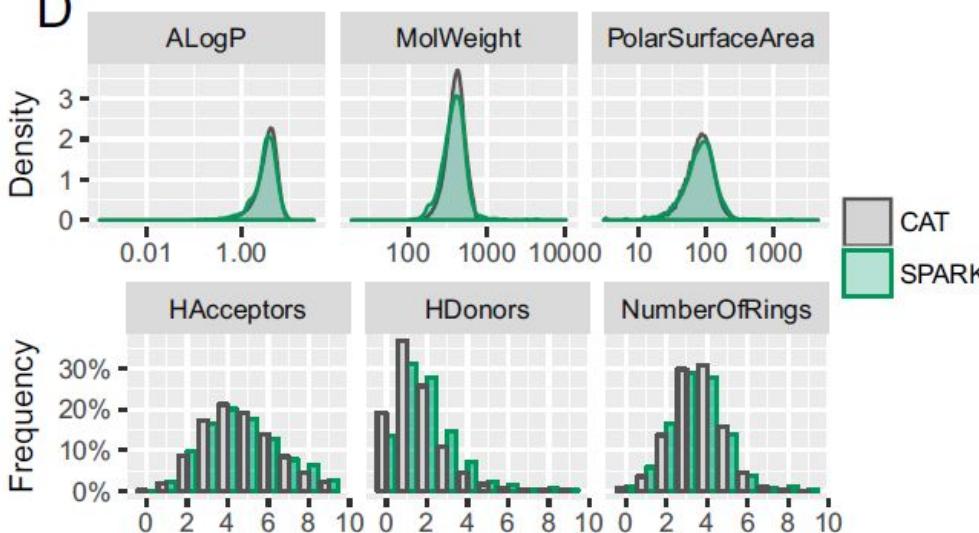
B



C



D

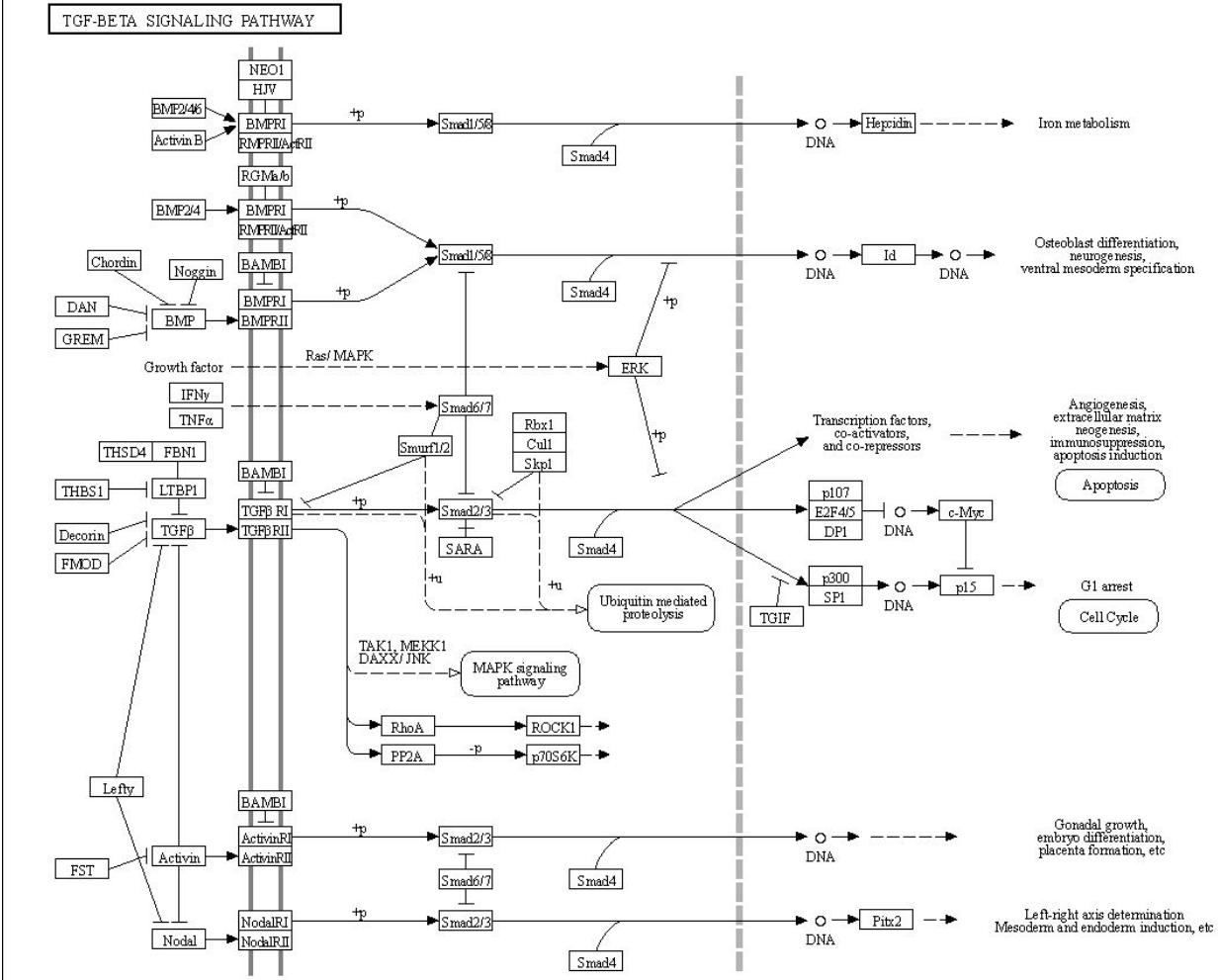


Roudnicky *et al.*, PNAS, 2020,  
<https://www.pnas.org/content/early/2020/08/04/1911532117>

# Mapping genes to biological pathways

Option 1: [KEGG pathways](#),  
with the example of [TGF-β signaling pathway](#).

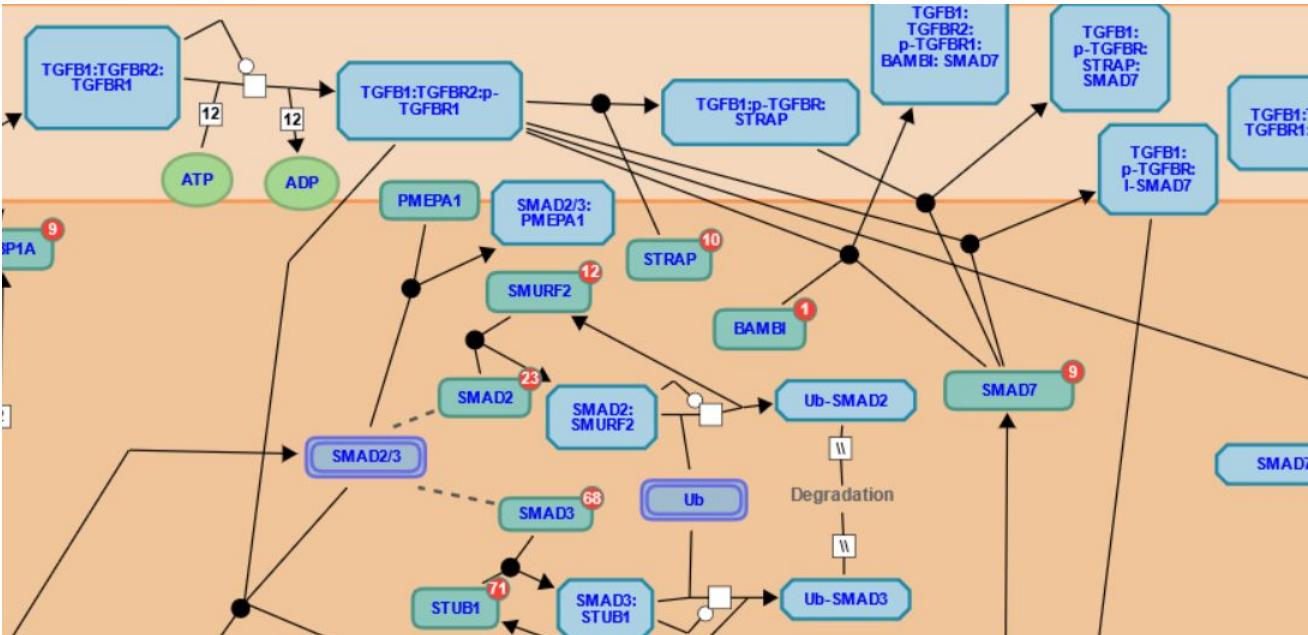
[A RESTful API](#) is available  
for academic use, with  
clients in Python and R.



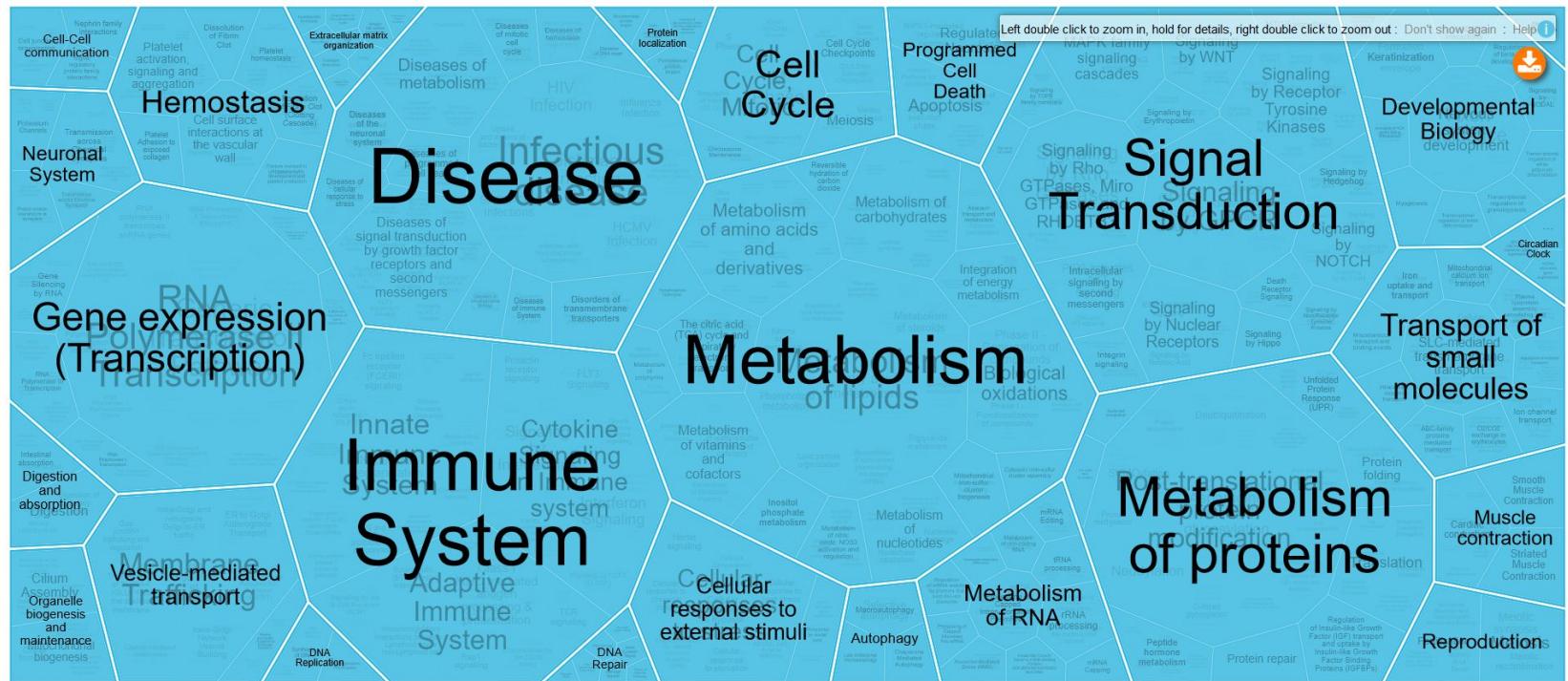
# Mapping genes to biological pathways

Option 2: [Reactome](#) pathways, with the example of the [TGF- \$\beta\$  signaling pathway](#).

[Developer's Zone](#) provides API and graph database interfaces.



# Overview of pathways captured by Reactome



The Voronoi (Reacfoam) view of all pathways in Reactome

# Mapping genes to biological processes

- Gene Ontology
- UniProtKB keywords
- Example:

## TGFBR2\_HUMAN

(TGF-beta receptor type -2, P37173)



### GO - Biological process<sup>i</sup>

- activation of protein kinase activity  Source: BHF-UCL
- aging  Source: Ensembl
- animal organ regeneration  Source: Ensembl
- apoptotic process  Source: UniProtKB
- atrioventricular valve morphogenesis  Source: BHF-UCL
- blood vessel development  Source: BHF-UCL
- brain development  Source: BHF-UCL

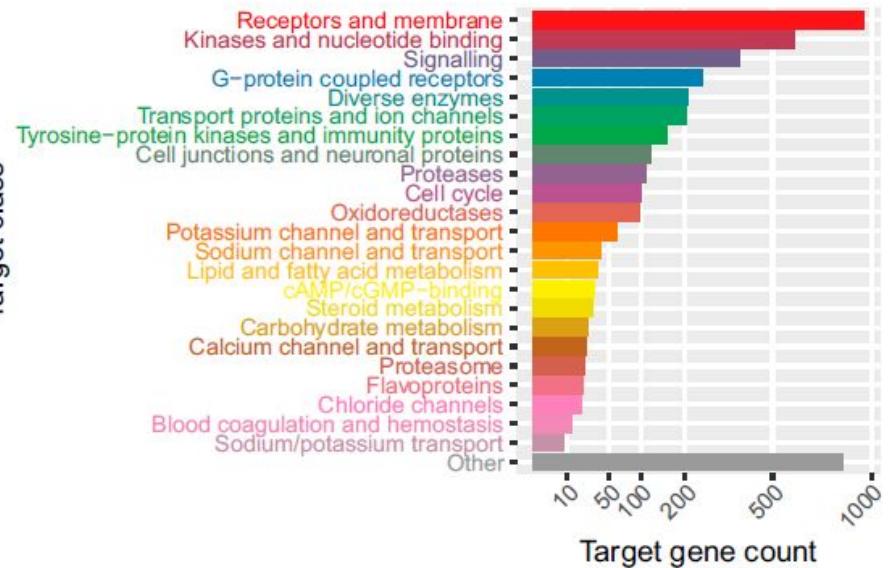


### Keywords<sup>i</sup>

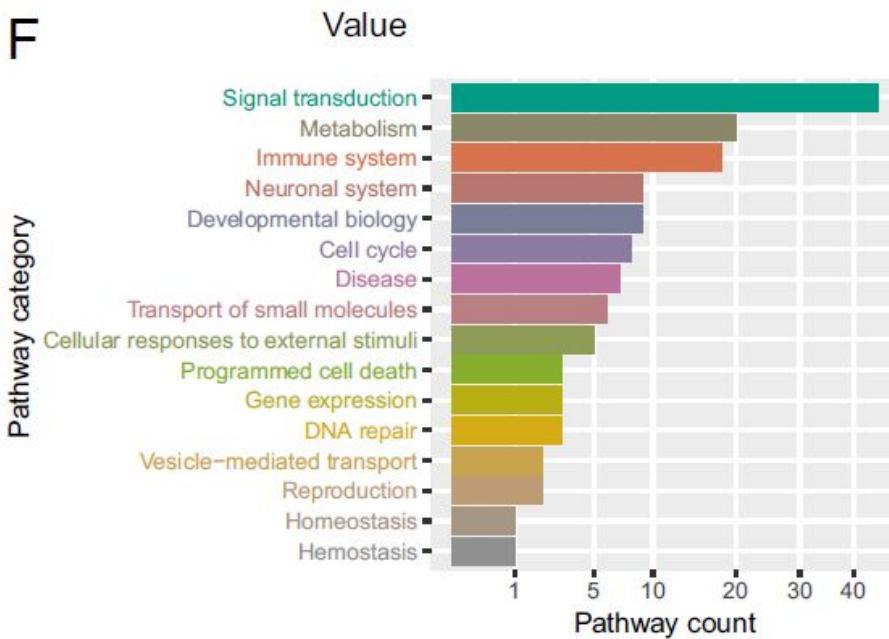
Molecular function	Kinase, Receptor, Serine/threonine-protein kinase, Transferase
Biological process	Apoptosis, Differentiation, Growth regulation
Ligand	ATP-binding, Magnesium, Manganese, Metal-binding, Nucleotide-binding

# SPARK covers the target space evenly with representative, potent, and specific compounds

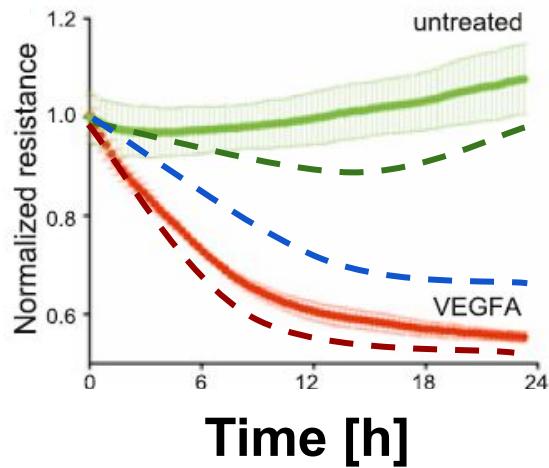
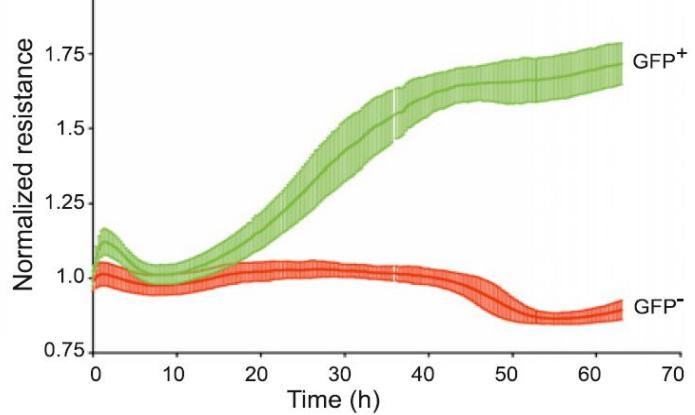
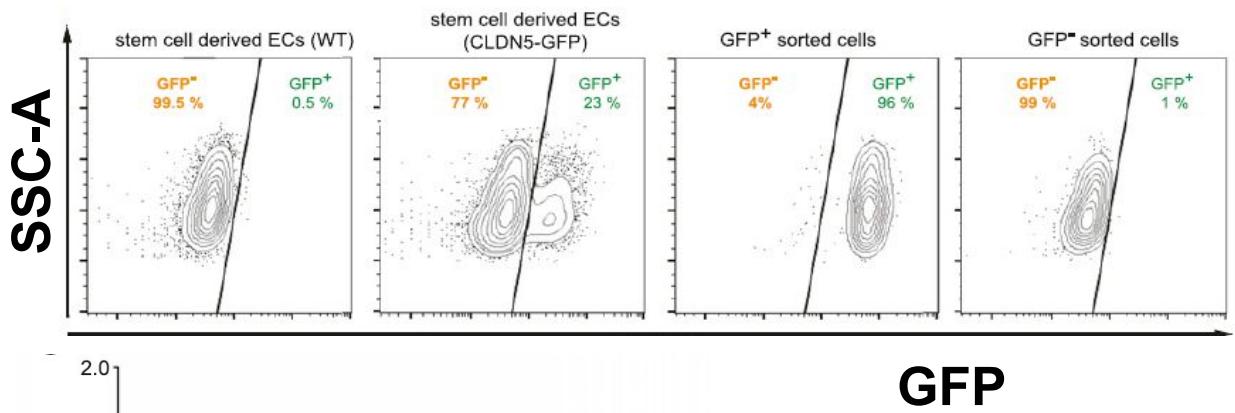
E



F

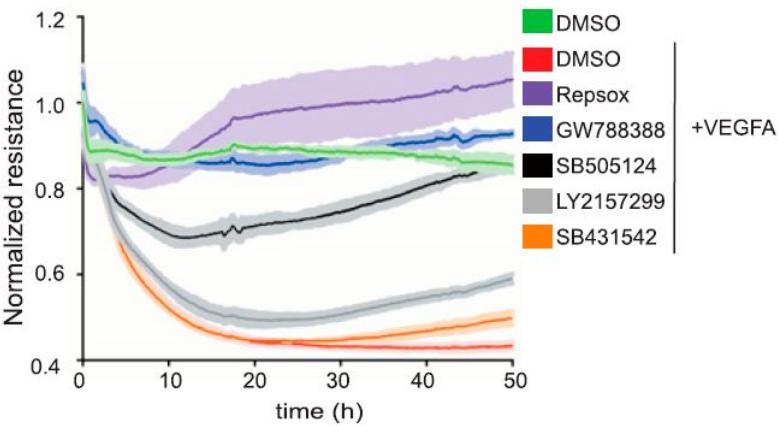
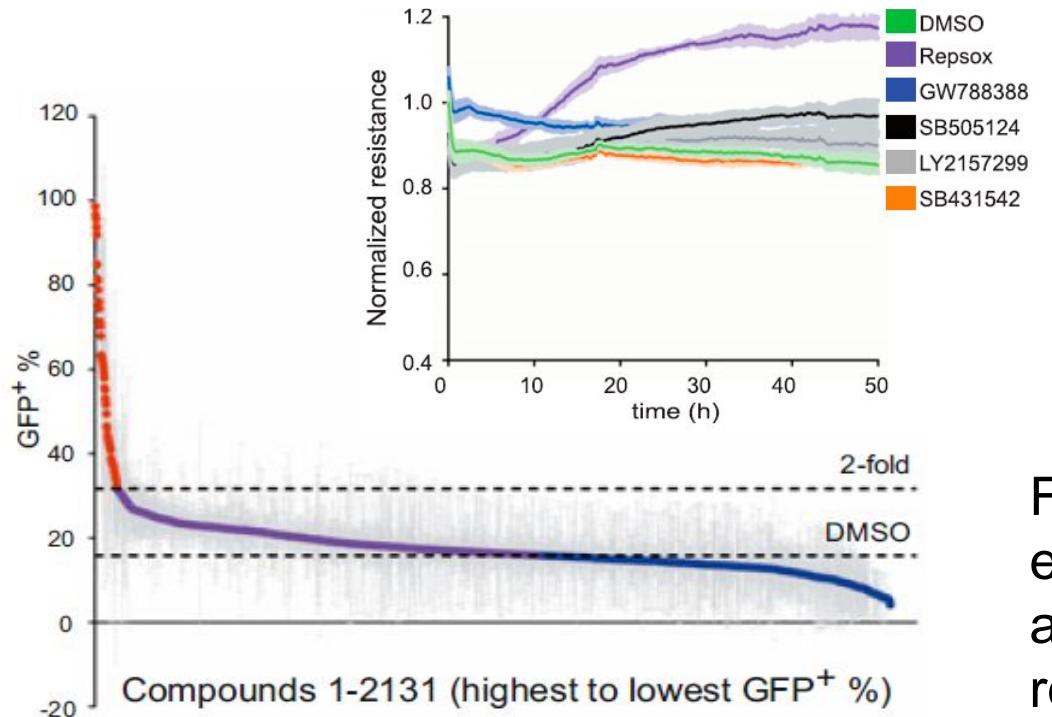


# Screening with stem-cell-derived endothelial cells with a reporter added by genome editing



SSC-A: Side-scatter area of flow cytometry;  
 GFP: Green fluorescent protein;  
 VEGFA: vascular endothelial growth factor A

# Compounds targeting the TGF- $\beta$ pathway such as RepSox modulates endothelial cells



Further *in vitro* and *in vivo* experiments establish RepSox as a tool compound modulating retinopathy.

# Do numbers of SNPs and conserved genome fit?

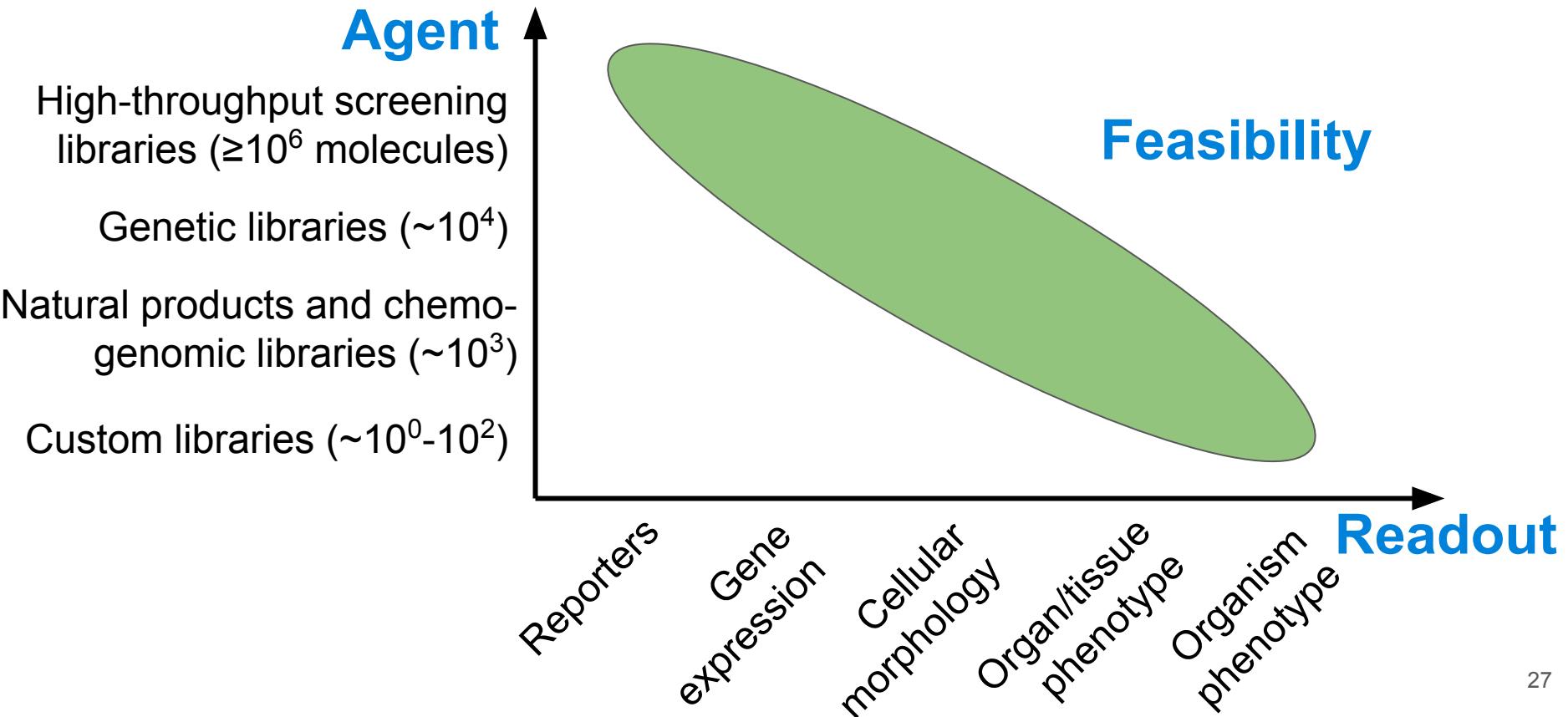
How can it be that individual genomes show only limited numbers of differences, while much of the genome is *junk*, i.e. not constrained by evolution, the loss or gain of which does not seriously affect fitness of the host organism?

## Facts:

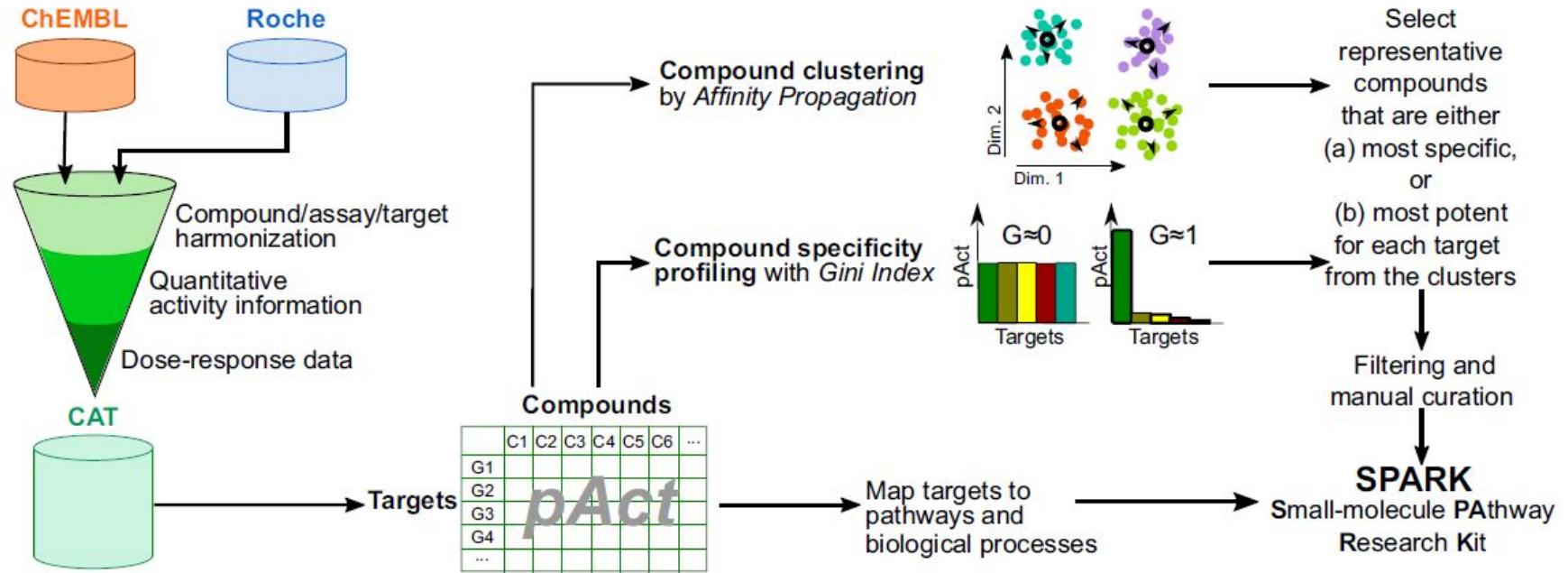
- On average, about 0.6% of the genome of an individual differs from the reference human genome (~5 million sites, affecting 20 million bases).
- Aggregating all known SNPs, we have detected 0.3 billion variants in sequenced samples.
- GWAS usually probes a subset of SNPs in linkage disequilibrium to identify causal variants. Therefore the intra-individual differences may be underestimated.

Numbers for <i>Homo sapiens</i>	Estimate	Source
History	~6.5E5 y	<a href="#">Timeline of the human condition</a>
Genome mutation rate	~1E-9/site/year	<a href="#">Lynch 2010, PNAS</a>
Genome size	~6E9 sites	<a href="#">Human genome hg38</a>
Estimated total mutations in the genome	~6.5E5 * 1E-9 * 6E9 * ~ 4E6, versus 5E6	<a href="#">Leypold 2021, Auton 2015</a>
<del>Effective population size</del>	<del>~1E4 individuals</del>	<a href="#">Charlesworth 2009, Park 2011</a>

# Phenotypic screenings by agent and readout



# Construction of SPARK in detail



## Harmonization

... of public and  
Roche internal data

## Machine learning

... to select  
compounds

## Pathways

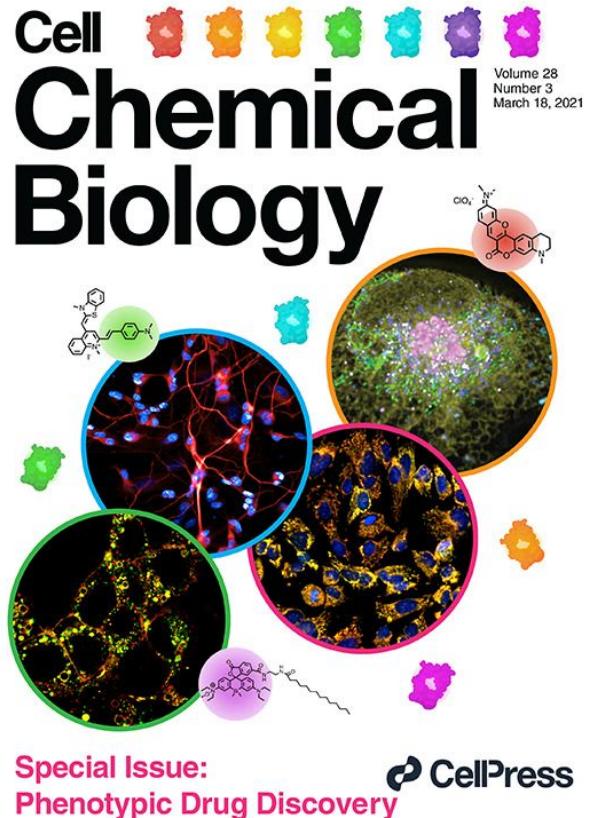
... mapped to  
compounds

## Curation

... to enrich quality  
compounds

# Conclusions about chemogenomic library

- Phenotypic drug discovery can lead to first-in-class drugs with novel mechanisms;
- Unsupervised machine learning and data modelling contribute to build chemogenomic libraries;
- We can link drug candidates via targets to biological pathways and processes.

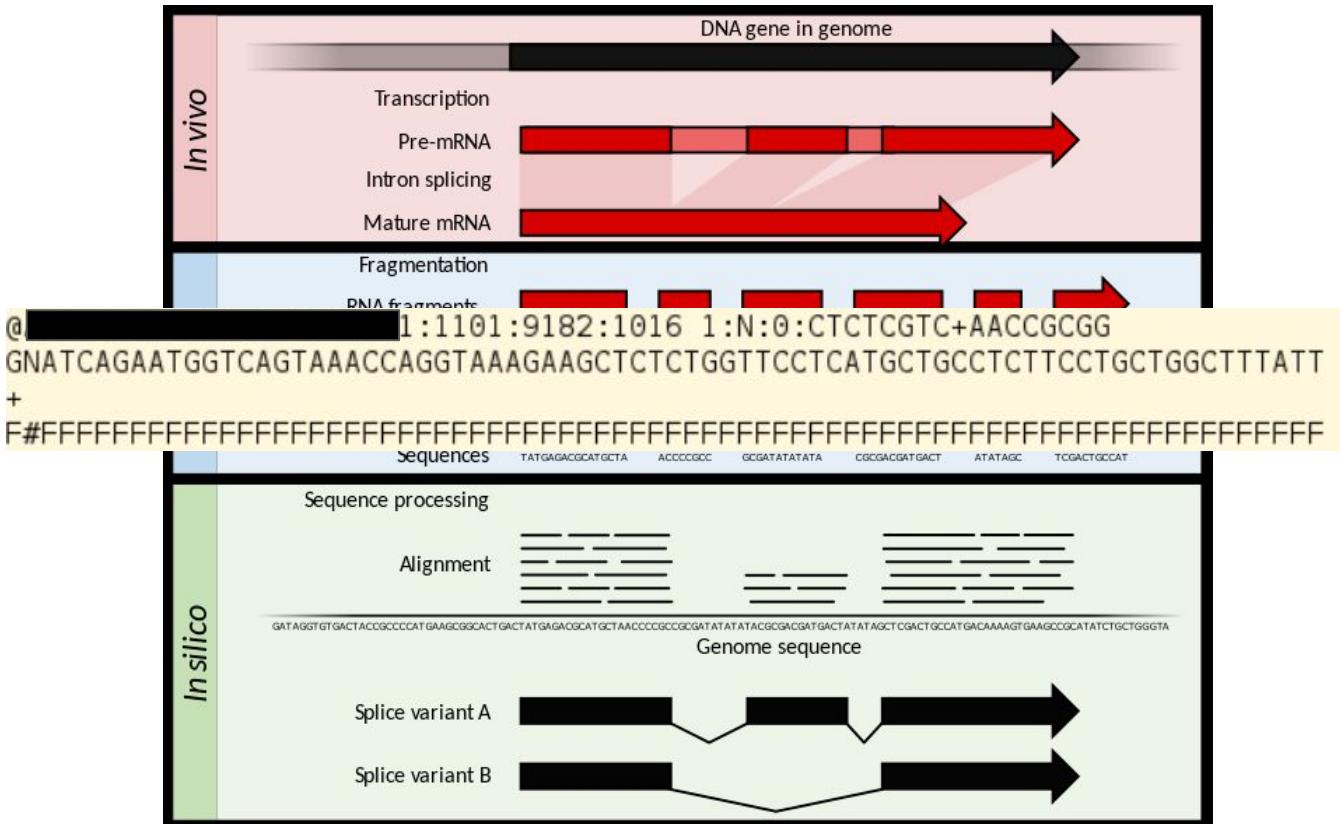


# Offline activities of Module II

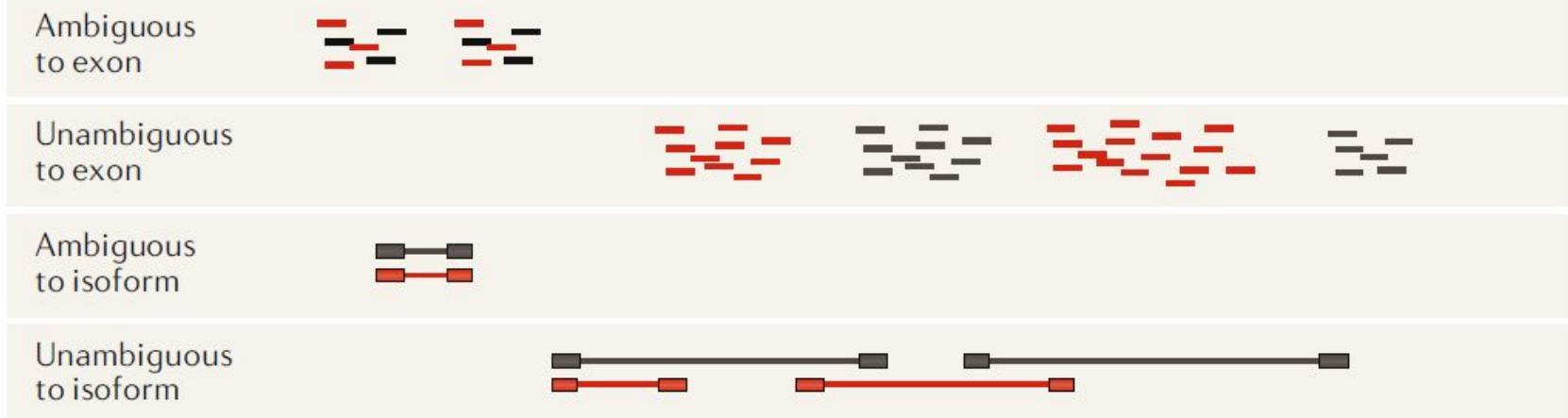
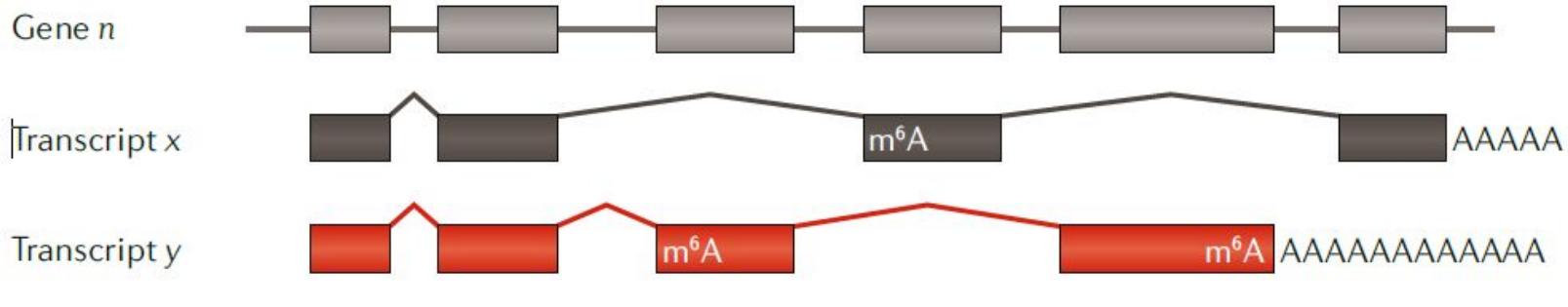
Please use your favourite programming language (shell scripts, python, R, for instance) and APIs (Application Programming Interfaces) of databases to perform following operations. Submit your code.

1. Retrieve all approved drugs from the ChEMBL database, sort them by approval year and name ([a Python example is here](#); documentations of the ChEMBL API can be found [here](#));
2. For each approved drug **since 2012** that you identified in step (1), retrieve a list of UniProt accession numbers, namely protein targets associated with the drug;
3. For each protein with a UniProt accession number that you identified in step (2), retrieve UniProt keywords associated with it. [You can use the UniProt API, documented here](#). [Python](#) and [R](#) clients are also available.

# Transcriptome profiling by RNA sequencing



# Transcriptome profiling by RNA sequencing



# Read Mapping

## Count collection

## Normalization by library size

## Differential Gene Expression Analysis



	sample A1	sample A2	sample B1	sample B2
gene 1	8	10	100	200
gene 2	14	15	15	40
gene 3	33	40	35	70
...	...	...	...	...
gene N	100	120	105	220

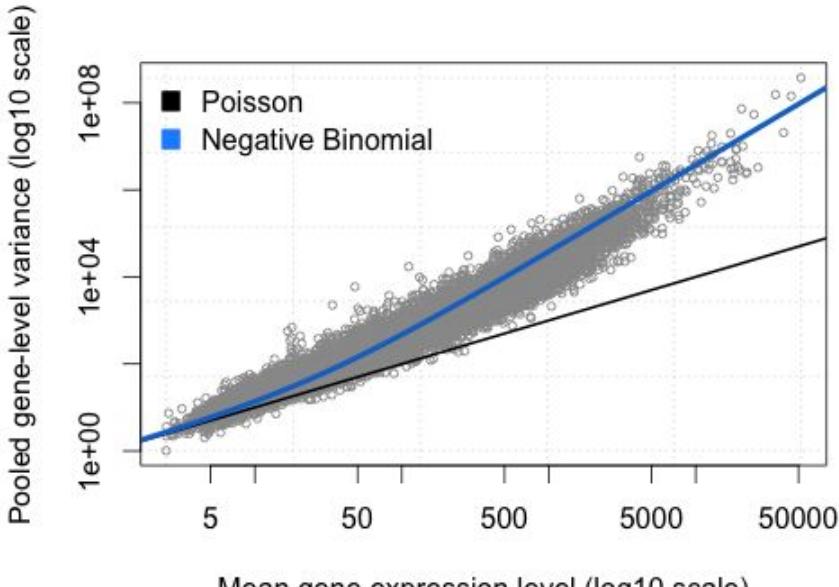
	sample A1	sample A2	sample B1	sample B2
gene 1	8	10	100	200
gene 2	14	15	115	40
gene 3	33	40	35	70
...	...	...	...	...
gene N	100	120	105	220

Tot. reads:  
5 millions

Tot. reads:  
10 millions

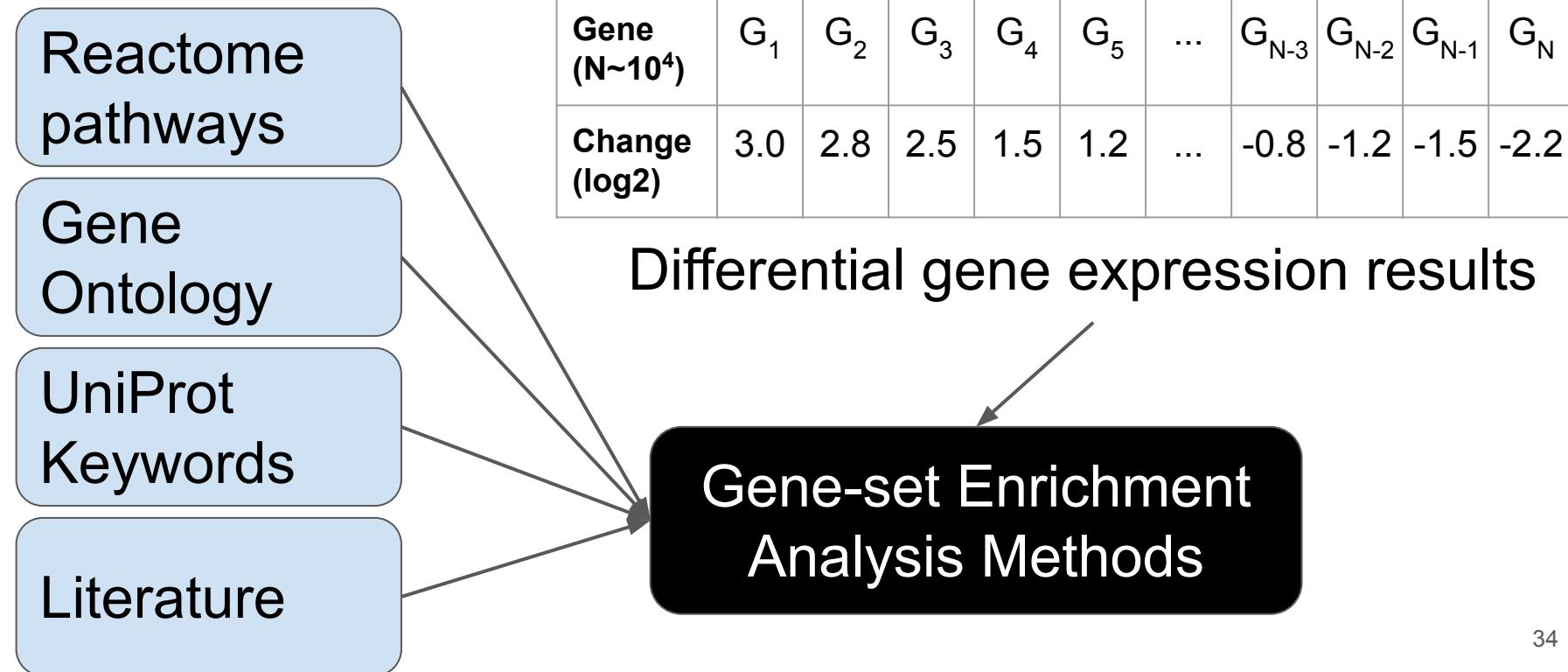
	sample A1	sample A2	sample B1	sample B2
gene 1	0.16	0.20	2.00	2.00
gene 2	0.28	0.30	0.30	0.40
gene 3	0.66	0.80	0.70	0.70
...	...	...	...	...
gene N	2.00	2.40	2.10	2.20

# Differential gene expression

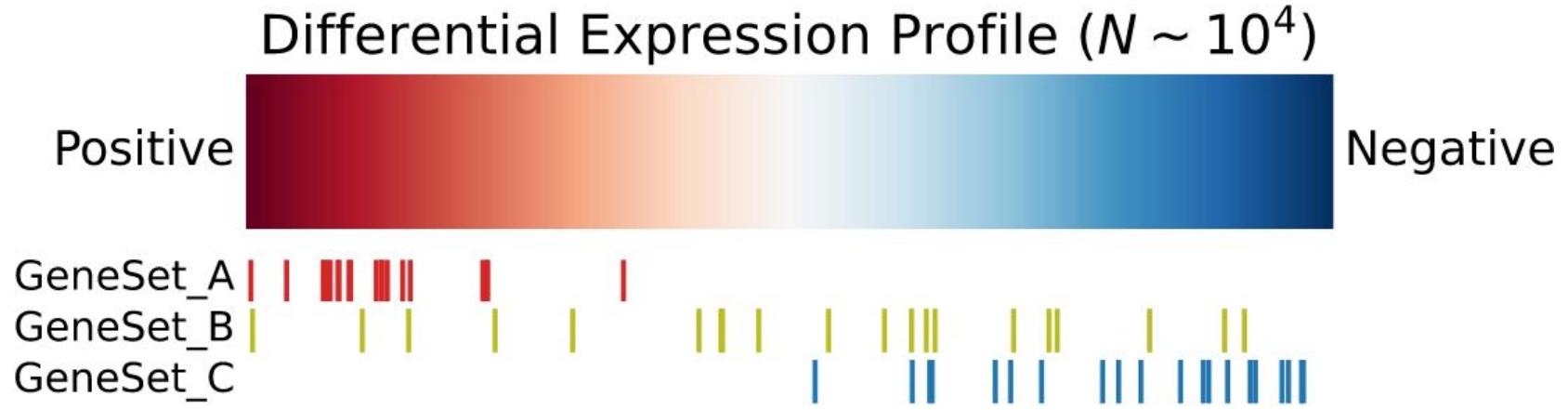


Tools: *edgeR* and *DESeq2*

# Interpret differential gene expression data with gene-set enrichment analysis



# Gene-set enrichment analysis



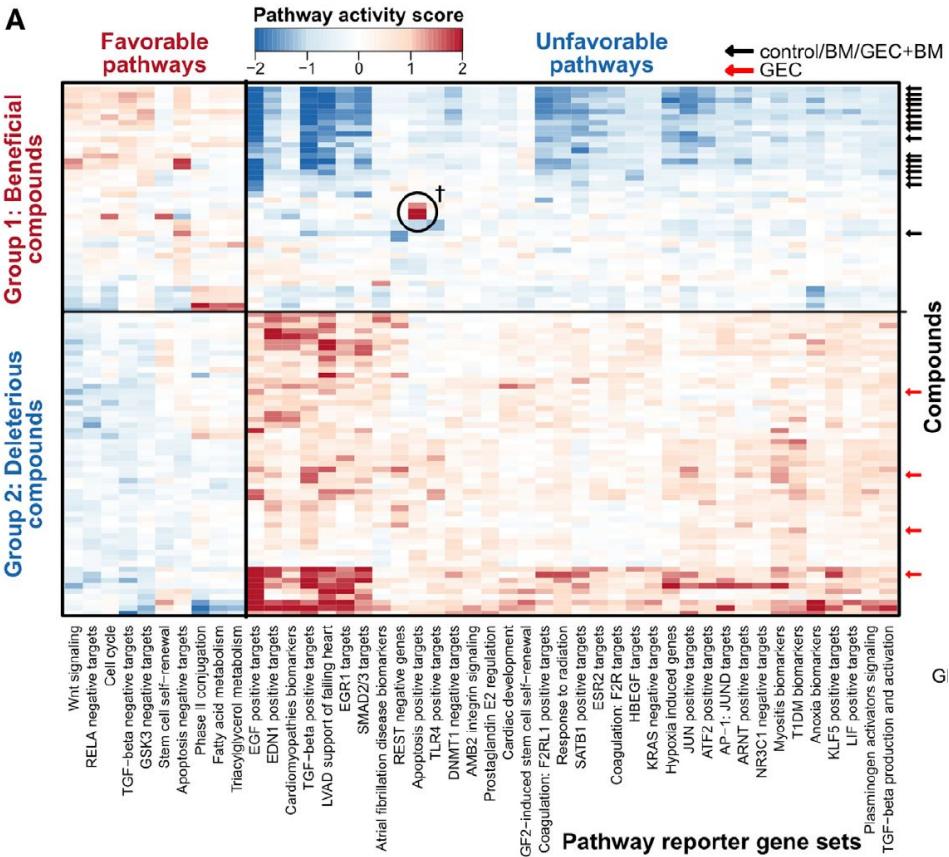
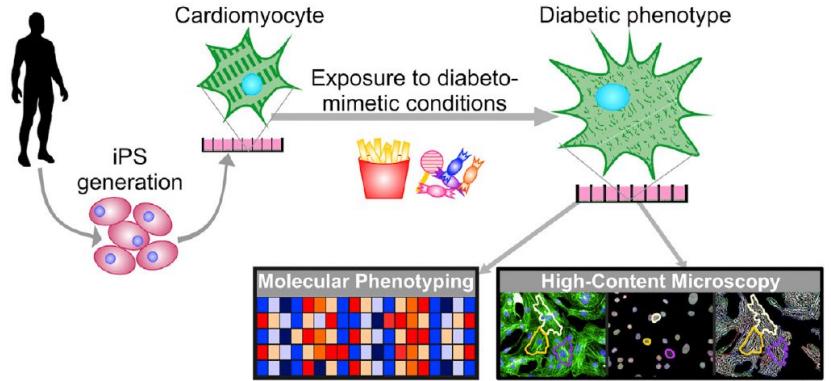
**Input:** (1) a differential gene expression profile; (2) a set of gene-sets  $\{G\}$ , each a set of genes.

**Output:** a ranked list of the input gene-sets by *enrichment*.

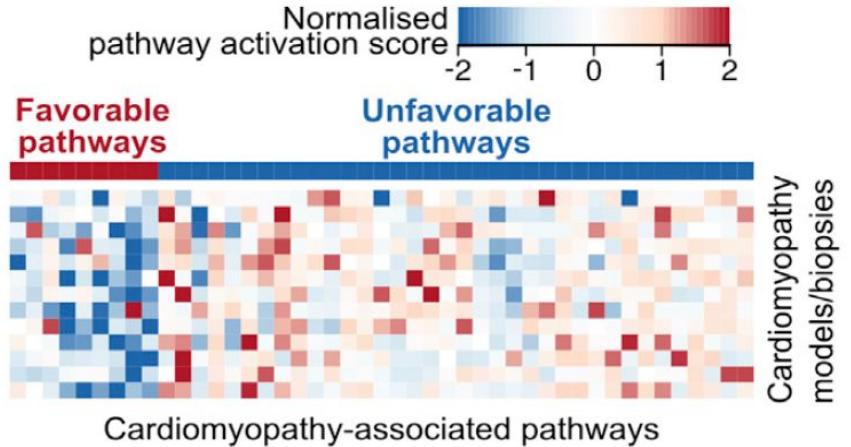
# Probability theory and statistical tools discussed

- Distributions
  - Gaussian distribution (used in linear model)
  - Bernoulli distribution → Binomial distribution → Negative binomial distribution
  - Poisson distribution → Negative binomial distribution
  - Poisson distribution  $\longleftrightarrow$  Exponential distribution
- Statistical methods
  - Bootstrapping method
  - Student's t-test
  - Wilcoxon-Mann-Whitney test
  - Kolmogorov-Smirnov test

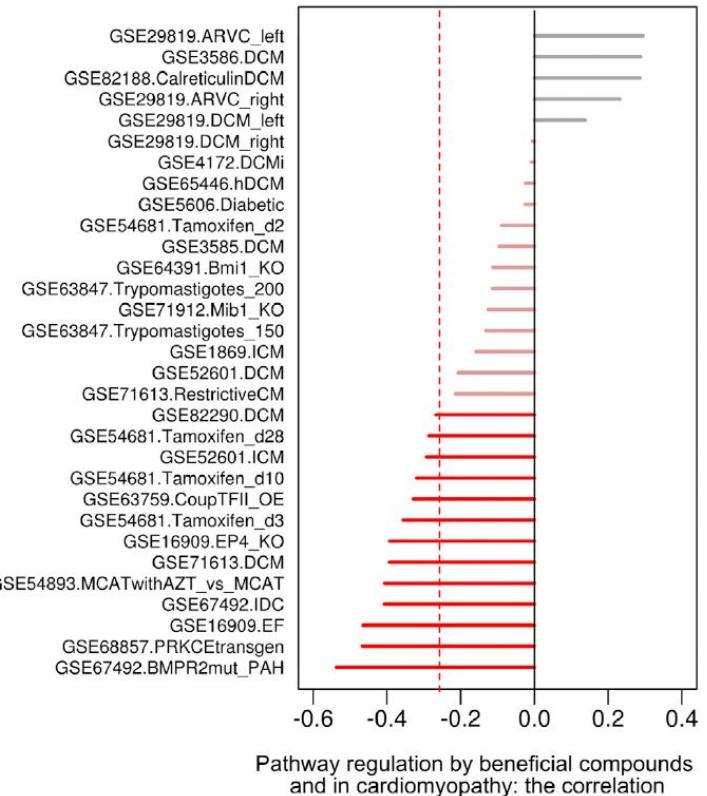
# Gene expression as screening readout



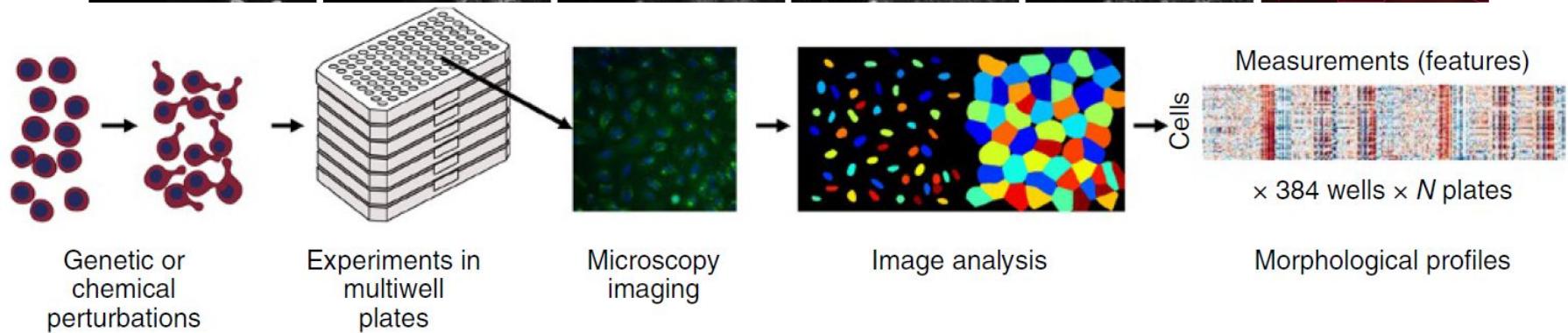
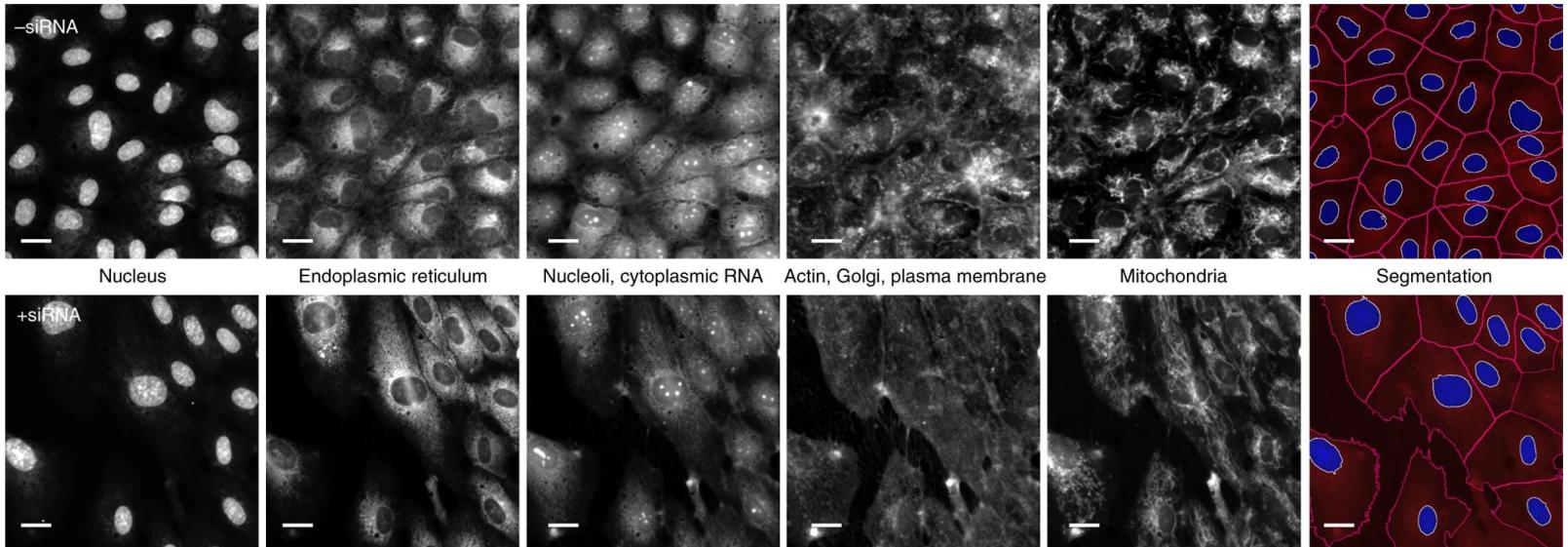
# Gene expression from patient and animal models help compound selection



We can prioritise molecules that reverse disease-induced changes.

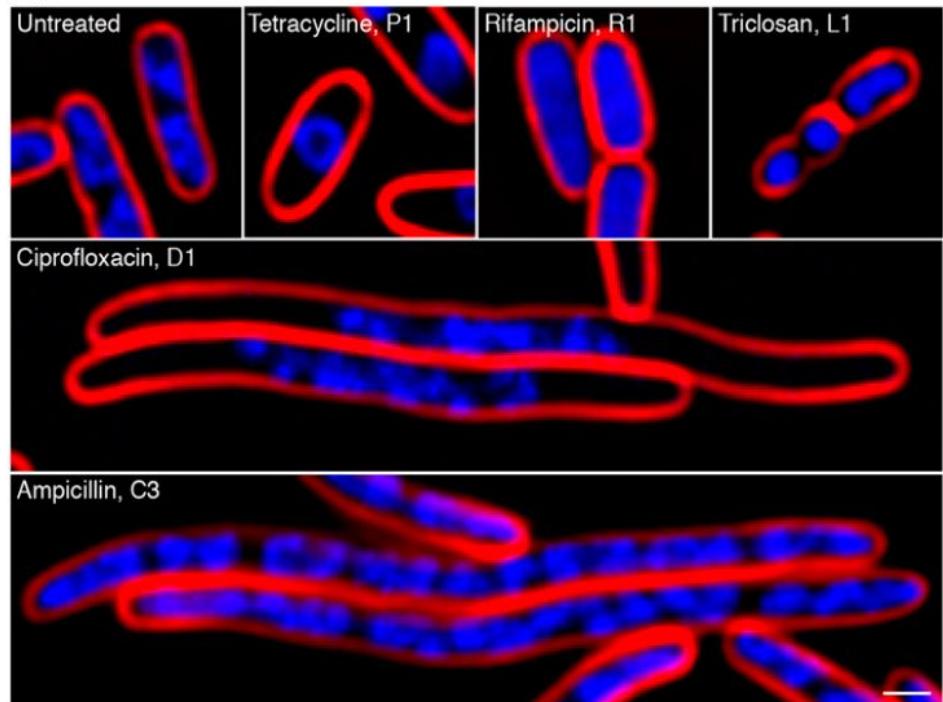


# Morphology as screening readout

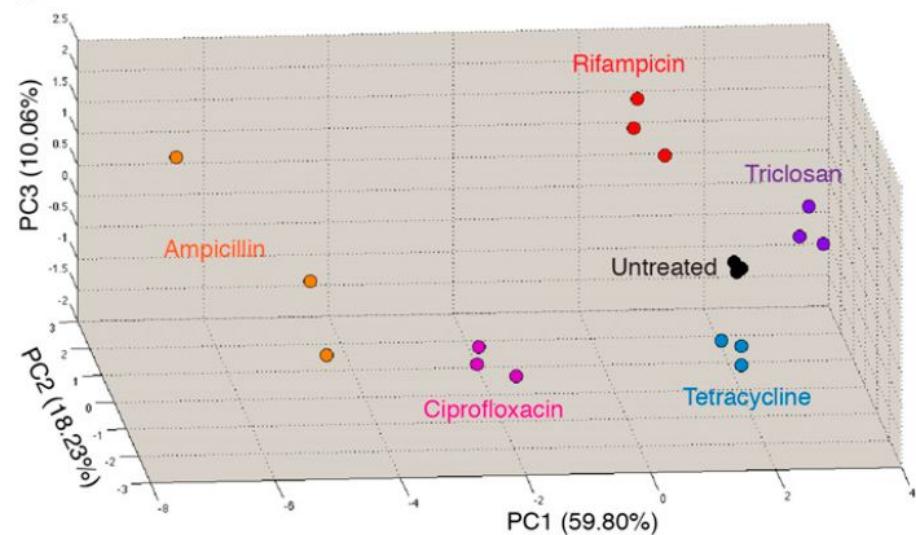


# Cytological profiling for antibiotics discovery

A



B



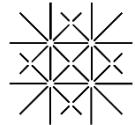
P: Protein translation inhibitors

R: RNA transcription inhibitors

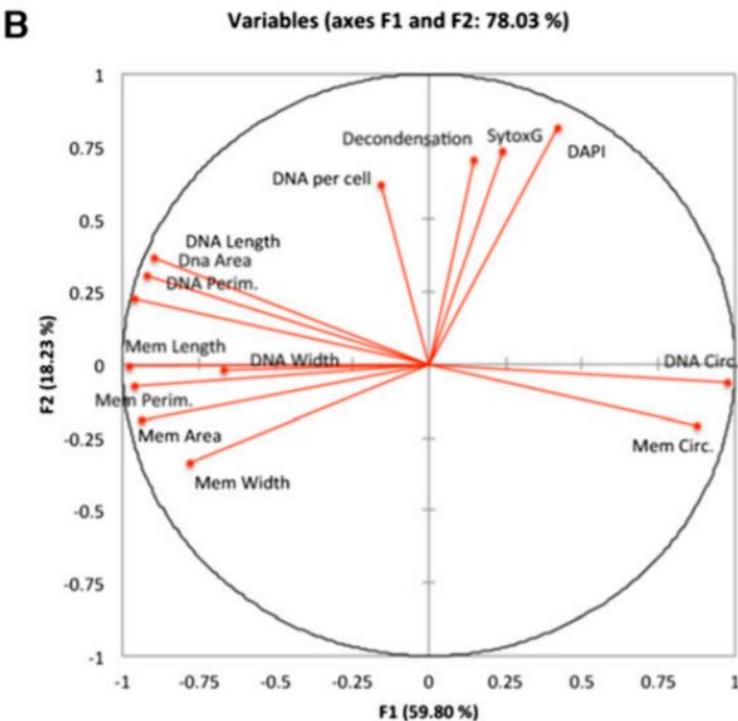
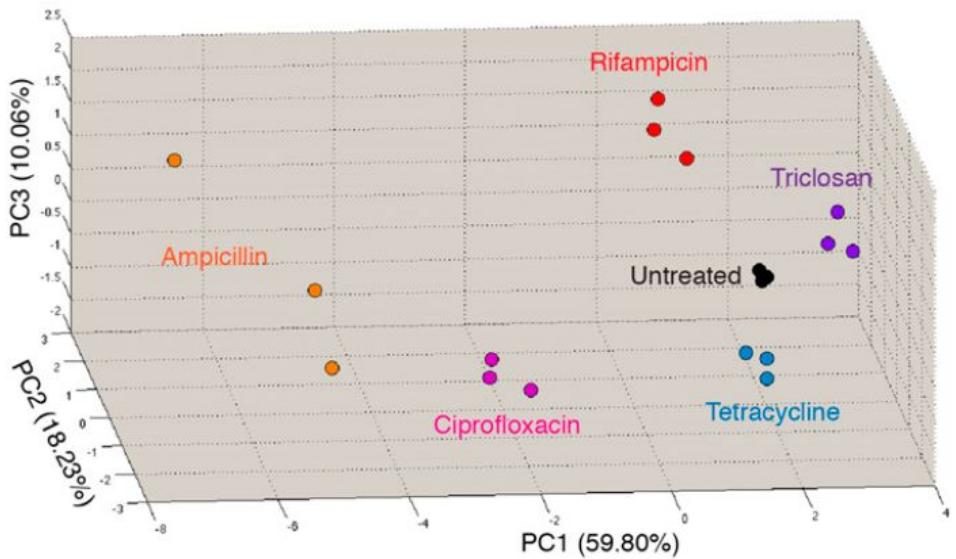
D: DNA replication inhibitors

L: Lipid biosynthesis inhibitors

C: Cell-wall synthesis inhibitors (peptidoglycan)



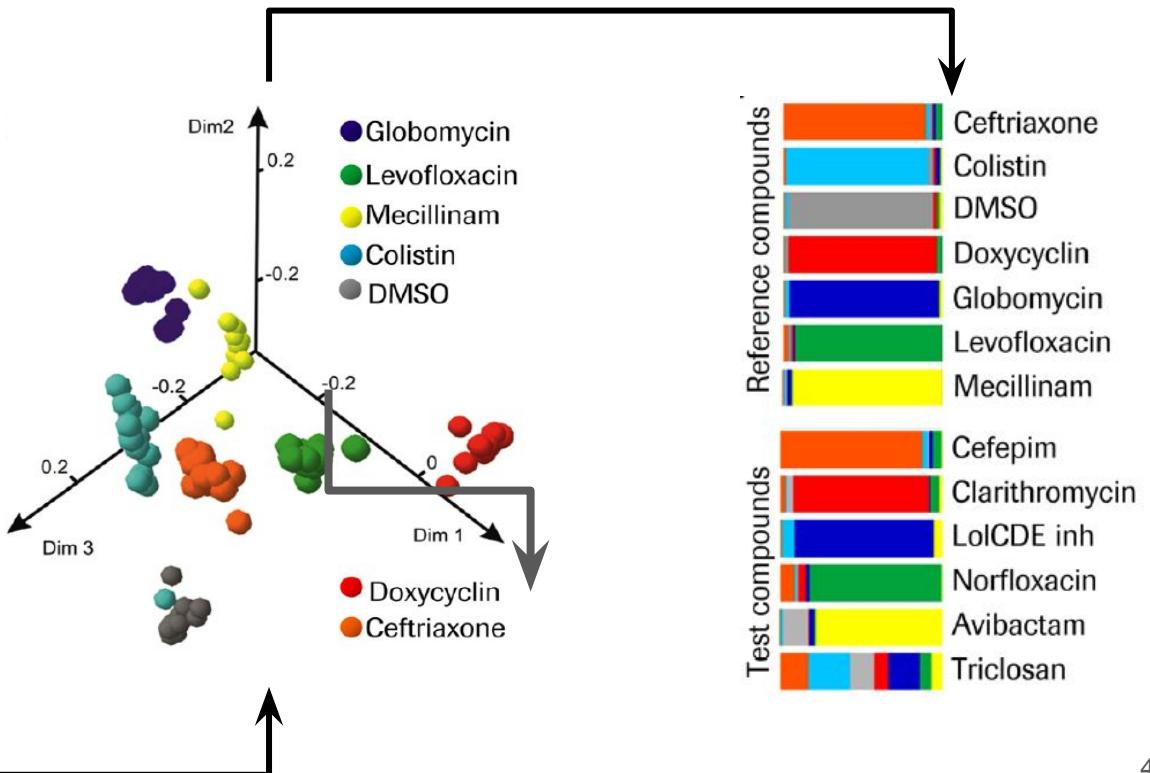
# Principal components are linear combination of morphological features



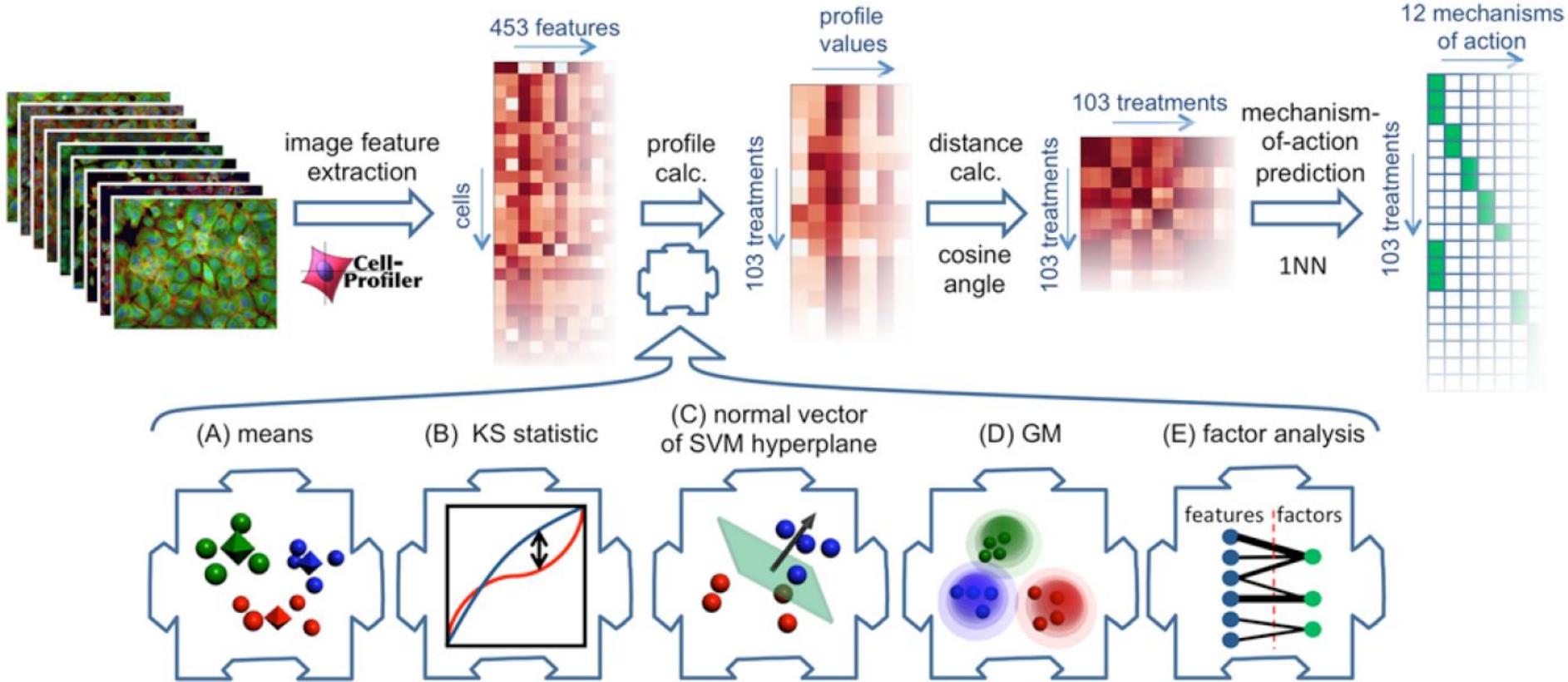
Membrane area, DNA area, Membrane perimeter, DNA perimeter, Membrane length, DNA length, No. of nucleoids per  
 $\mu\text{m}^2$   $\mu\text{m}^2$   $\mu\text{m}$   $\mu\text{m}$   $\mu\text{m}$  cell

Membrane width, DNA width,  
 $\mu\text{m}$   $\mu\text{m}$  Membrane circularity DNA circularity SytoxG intensity DAPI intensity Decondensation

# Morphology classifies compounds by MoA



# Comparison of computational methods



# Do the benchmark and use Occam's Razor

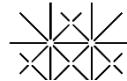
**Table I.** Accuracies for classifying compound treatments into mechanisms of action.

Method	Accuracy, %
Means	83
KS statistic	83
Normal vector to support-vector machine hyperplane	81
With recursive feature elimination	64
Distribution over Gaussian mixture components	83
Factor analysis + means	94

True mechanistic class

	Act	Aur	Ch	DD	DR	Eg5	Epi	KI	MD	MS	PD	PS	Acc.
Act	4												80 %
Aur		12											100 %
Ch			6										100 %
DD				8									93 %
DR					1	7							81 %
Eg5						12							100 %
Epi							2						76 %
KI								6					100 %
MD									5				100 %
MS										13			93 %
PD											9		97 %
PS												7	96 %
												8	100 %

Overall accuracy: 94 %

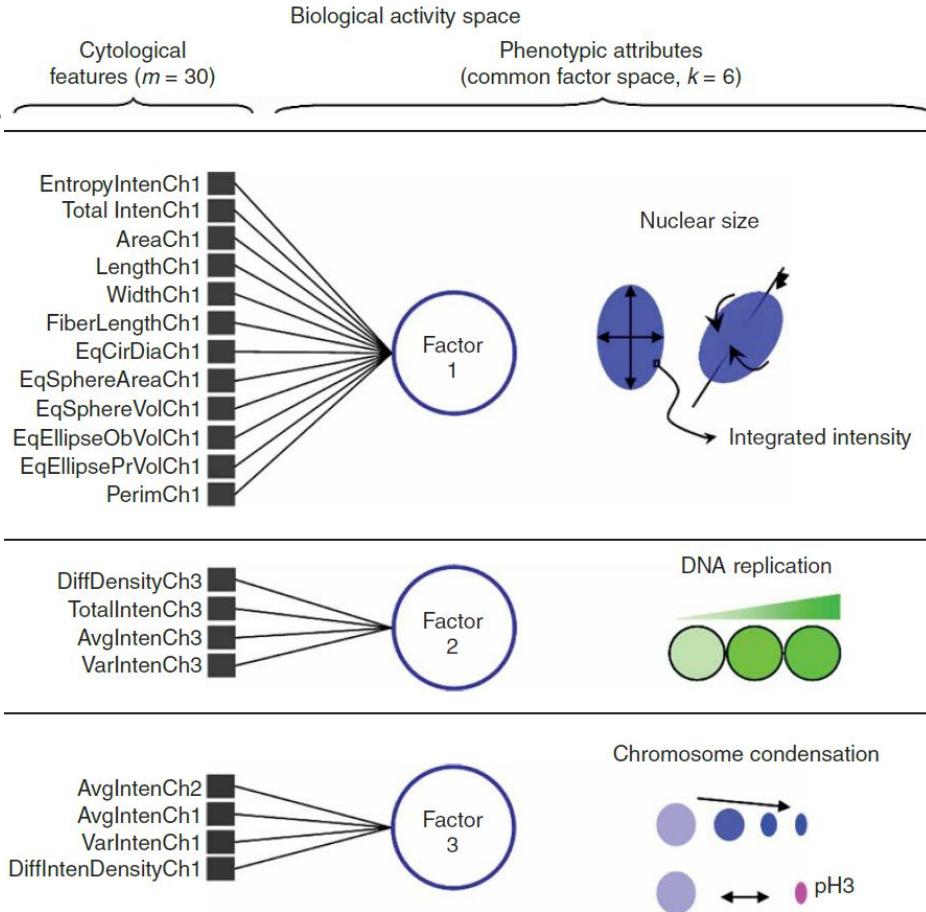


# A possible explanation for the success of latent variable models

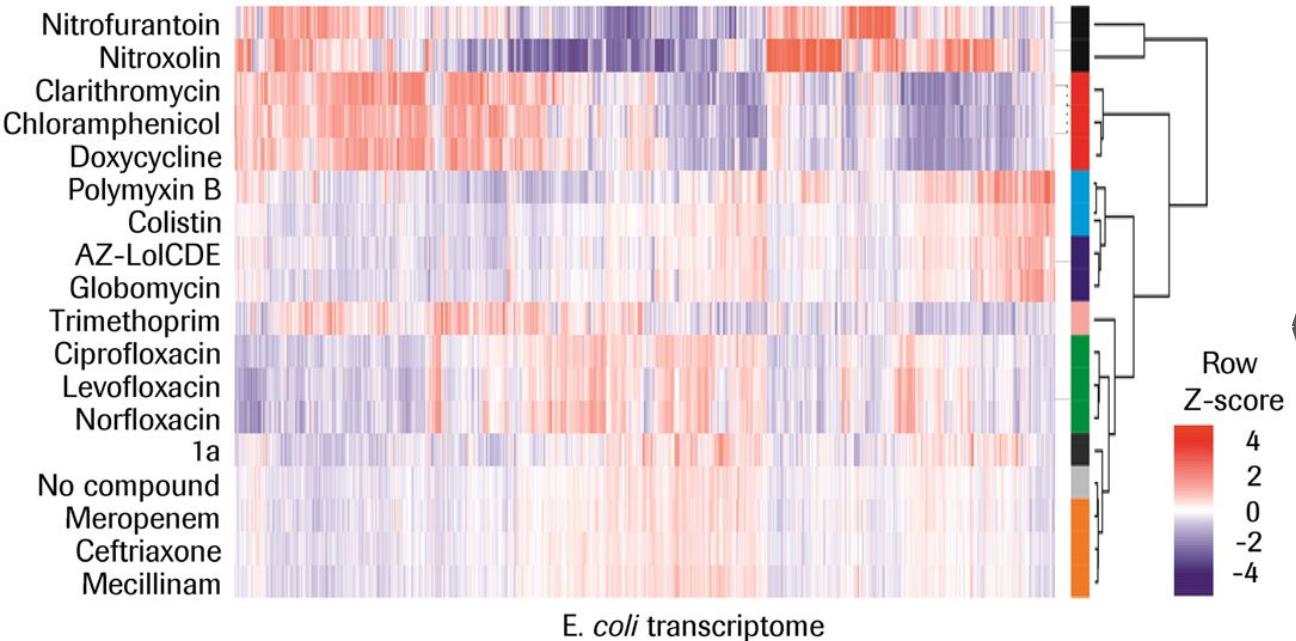
$$\text{Cells} \begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{nl} & \cdots & x_{nm} \end{pmatrix} = X_{nm} = \sum_{i=1}^k L_{ni} F_{im} + \varepsilon_{nm}$$

Cytological features    k-factor space

A common latent factor model



# Morphology and gene expression used jointly



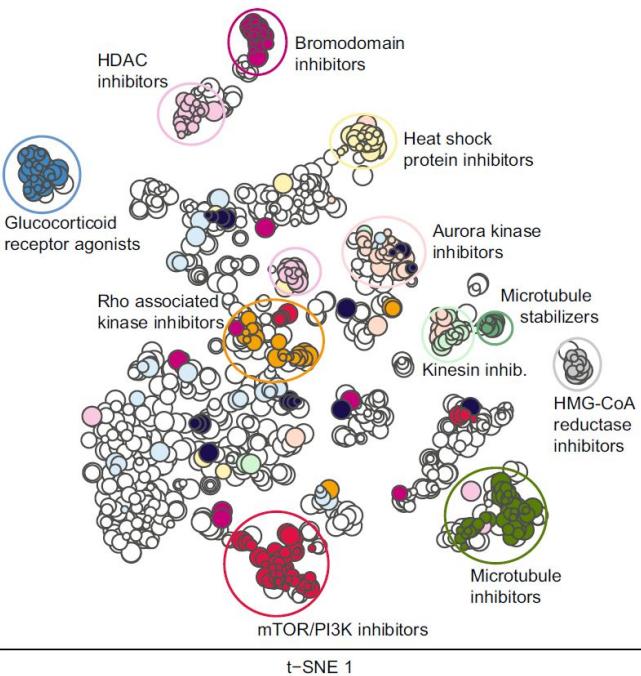
Gene-set enrichment analysis

Reporter assays

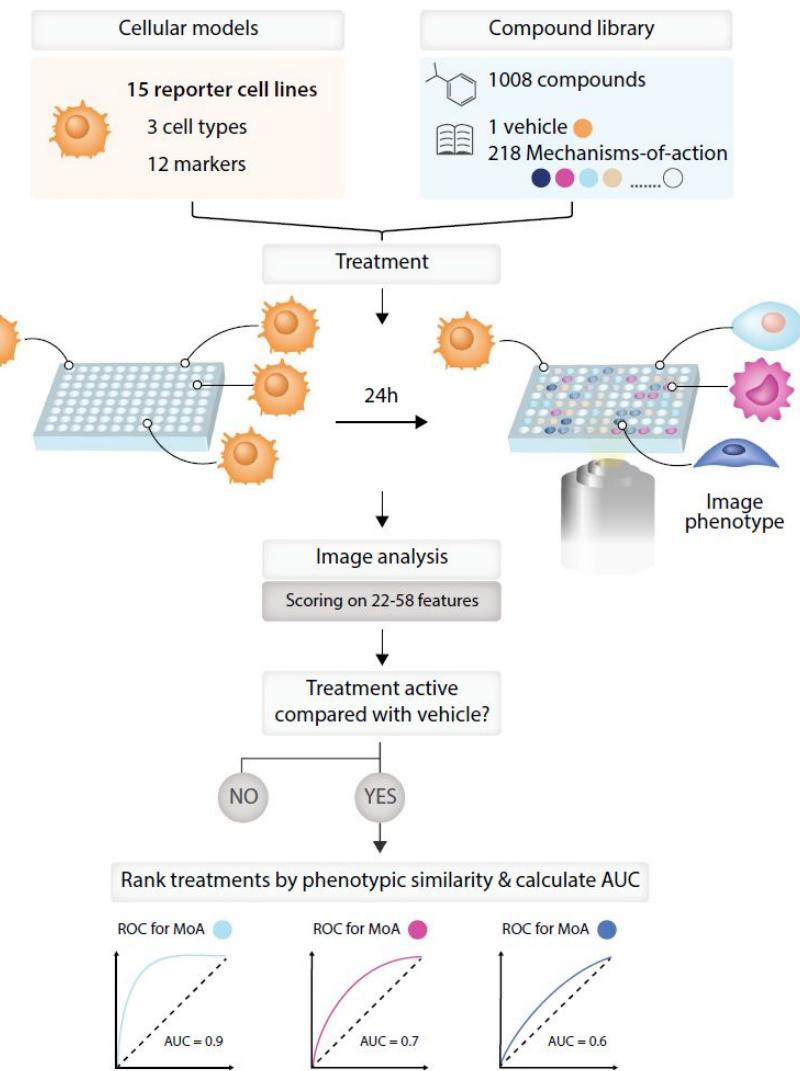
Pathway-  
Phenotype associations

# A multi-cell-type, 1008-compound screening by Cox et al. (2020)

t-SNE 2



- ABL1 inhibitor
- Aurora kinase inhibitor
- Bromodomain inhibitor
- Glucocorticoid receptor agonist
- HDAC inhibitor
- Heat shock protein inhibitor
- HMG-CoA reductase inhibitor
- Kinesin inhibitor
- Microtubule inhibitor
- Microtubule stabilizer
- mTOR and/or PI3K inhibitor
- Rho associated kinase inhibitor
- VEGFR family inhibitor
- (Other)



# Conclusions

- Gene expression and image-based profiling can be used individually or jointly for phenotypic screening;
- Integration of biological knowledge, high-throughput data, and statistical modelling empowers phenotypic drug discovery.

# References

1. Swinney, David C., and Jason Anthony. 2011. "How Were New Medicines Discovered?" *Nature Reviews Drug Discovery* 10 (7): 507–19. <https://doi.org/10.1038/nrd3480>.
2. Marx, Uwe, Tommy B. Andersson, Anthony Bahinski, Mario Beilmann, Sonja Beken, Flemming R. Cassee, Murat Cirit, et al. 2016. "Biology-Inspired Microphysiological System Approaches to Solve the Prediction Dilemma of Substance Testing." *ALTEX - Alternatives to Animal Experimentation* 33 (3): 272–321. <https://doi.org/10.14573/altex.1603161>.
3. Dickey, Seth W., Gordon Y. C. Cheung, and Michael Otto. 2017. "Different Drugs for Bad Bugs: Antivirulence Strategies in the Age of Antibiotic Resistance." *Nature Reviews Drug Discovery* 16 (7): 457–71. <https://doi.org/10.1038/nrd.2017.23>.
4. Lewis, Kim. 2013. "Platforms for Antibiotic Discovery." *Nature Reviews Drug Discovery* 12 (5): 371–87. <https://doi.org/10.1038/nrd3975>.
5. Warner, Katherine Deigan, Christine E. Hajdin, and Kevin M. Weeks. 2018. "Principles for Targeting RNA with Drug-like Small Molecules." *Nature Reviews Drug Discovery* 17 (8): 547–58. <https://doi.org/10.1038/nrd.2018.93>.
6. Berg, Ellen L. 2021. "The Future of Phenotypic Drug Discovery." *Cell Chemical Biology* 28 (3): 424–30. <https://doi.org/10.1016/j.chembiol.2021.01.010>.
7. Chin, Marcus Y., Jether Amos Espinosa, Grace Pohan, Sarine Markossian, and Michelle R. Arkin. 2021. "Reimagining Dots and Dashes: Visualizing Structure and Function of Organelles for High-Content Imaging Analysis." *Cell Chemical Biology* 28 (3): 320–37. <https://doi.org/10.1016/j.chembiol.2021.01.016>.
8. Conway, Louis P., Weichao Li, and Christopher G. Parker. 2021. "Chemoproteomic-Enabled Phenotypic Screening." *Cell Chemical Biology* 28 (3): 371–93. <https://doi.org/10.1016/j.chembiol.2021.01.012>.
9. Dahlin, Jayme L., Douglas S. Auld, Ina Rothenaigner, Steve Haney, Jonathan Z. Sexton, J. Willem M. Nissink, Jarrod Walsh, et al. 2021. "Nuisance Compounds in Cellular Assays." *Cell Chemical Biology* 28 (3): 356–70. <https://doi.org/10.1016/j.chembiol.2021.01.021>.
10. Ha, Jaeyoung, Hankum Park, Jongmin Park, and Seung Bum Park. 2021. "Recent Advances in Identifying Protein Targets in Drug Discovery." *Cell Chemical Biology* 28 (3): 394–423. <https://doi.org/10.1016/j.chembiol.2020.12.001>.

# References (continued)

11. Hsu, Ku-Lung. 2021. "Shining a Light on Phenotypic Drug Discovery." *Cell Chemical Biology* 28 (2): 115–17. <https://doi.org/10.1016/j.chembiol.2021.01.020>.
12. Hughes, Rebecca E., Richard J. R. Elliott, John C. Dawson, and Neil O. Carragher. 2021. "High-Content Phenotypic and Pathway Profiling to Advance Drug Discovery in Diseases of Unmet Need." *Cell Chemical Biology* 28 (3): 338–55. <https://doi.org/10.1016/j.chembiol.2021.02.015>.
13. Vandana, J. Jeya, Lauretta A. Lacko, and Shuibing Chen. 2021. "Phenotypic Technologies in Stem Cell Biology." *Cell Chemical Biology* 28 (3): 257–70. <https://doi.org/10.1016/j.chembiol.2021.02.001>.
14. Ziegler, Slava, Sonja Sievers, and Herbert Waldmann. 2021. "Morphological Profiling of Small Molecules." *Cell Chemical Biology* 28 (3): 300–319. <https://doi.org/10.1016/j.chembiol.2021.02.012>.
15. Moffat, John G., Fabien Vincent, Jonathan A. Lee, Jörg Eder, and Marco Prunotto. 2017. "Opportunities and Challenges in Phenotypic Drug Discovery: An Industry Perspective." *Nature Reviews Drug Discovery* 16 (8): 531–43. <https://doi.org/10.1038/nrd.2017.111>.
16. Drawnel, Faye Marie, Jitao David Zhang, Erich Küng, Natsuyo Aoyama, Fethallah Benmansour, Andrea Araujo Del Rosario, Sannah Jensen Zoffmann, et al. 2017. "Molecular Phenotyping Combines Molecular Information, Biological Relevance, and Patient Data to Improve Productivity of Early Drug Discovery." *Cell Chemical Biology* 18 (24(5)): 624–34. <https://doi.org/10.1016/j.chembiol.2017.03.016>.
17. Roudnicki, Filip, Jitao David Zhang, Bo Kyoung Kim, Nikhil J. Pandya, Yanjun Lan, Lisa Sach-Peltason, Heloise Ragelle, et al. 2020. "Inducers of the Endothelial Cell Barrier Identified through Chemogenomic Screening in Genome-Edited HPSC-Endothelial Cells." *Proceedings of the National Academy of Sciences*, August. <https://doi.org/10.1073/pnas.1911532117>.
18. Zoffmann, Sannah, Maarten Vercruyse, Fethallah Benmansour, Andreas Maunz, Luise Wolf, Rita Blum Marti, Tobias Heckel, et al. 2019. "Machine Learning-Powered Antibiotics Phenotypic Drug Discovery." *Scientific Reports* 9 (1): 1–14. <https://doi.org/10.1038/s41598-019-39387-9>.
19. Nichols, Robert J., Saunak Sen, Yoe Jin Choo, Pedro Beltrao, Matylda Zietek, Rachna Chaba, Sueyoung Lee, et al. 2011. "Phenotypic Landscape of a Bacterial Cell." *Cell* 144 (1): 143–56. <https://doi.org/10.1016/j.cell.2010.11.052>.
20. Smith, Kevin, Filippo Piccinini, Tamas Balassa, Krisztian Koos, Tivadar Danka, Hossein Azizpour, and Peter Horvath. 2018. "Phenotypic Image Analysis Software Tools for Exploring and Understanding Big Image Data from Cell-Based Assays." *Cell Systems* 6 (6): 636–53. <https://doi.org/10.1016/j.cels.2018.06.001>.

# References (continued)

21. Scheeder, Christian, Florian Heigwer, and Michael Boutros. 2018. "Machine Learning and Image-Based Profiling in Drug Discovery." *Current Opinion in Systems Biology, Pharmacology and drug discovery*, 10 (August): 43–52. <https://doi.org/10.1016/j.coisb.2018.05.004>.
22. Schirle, Markus, and Jeremy L. Jenkins. 2016. "Identifying Compound Efficacy Targets in Phenotypic Drug Discovery." *Drug Discovery Today* 21 (1): 82–89. <https://doi.org/10.1016/j.drudis.2015.08.001>.
23. Wilkinson, Isabel V. L., Georg C. Terstappen, and Angela J. Russell. 2020. "Combining Experimental Strategies for Successful Target Deconvolution." *Drug Discovery Today* 25 (11): 1998–2005. <https://doi.org/10.1016/j.drudis.2020.09.016>.
24. Comess, Kenneth M., Shaun M. McLoughlin, Jon A. Oyer, Paul L. Richardson, Henning Stöckmann, Anil Vasudevan, and Scott E. Warder. 2018. "Emerging Approaches for the Identification of Protein Targets of Small Molecules - A Practitioners' Perspective." *Journal of Medicinal Chemistry* 61 (19): 8504–35. <https://doi.org/10.1021/acs.jmedchem.7b01921>.
25. Sidders, Ben, Anna Karlsson, Linda Kitching, Rubben Torella, Paul Karila, and Anne Phelan. 2018. "Network-Based Drug Discovery: Coupling Network Pharmacology with Phenotypic Screening for Neuronal Excitability." *Journal of Molecular Biology, Theory and Application of Network Biology Toward Precision Medicine*, 430 (18, Part A): 3005–15. <https://doi.org/10.1016/j.jmb.2018.07.016>.
26. Aulner, Nathalie, Anne Danckaert, JongEun Ihm, David Shum, and Spencer L. Shorte. 2019. "Next-Generation Phenotypic Screening in Early Drug Discovery for Infectious Diseases." *Trends in Parasitology* 35 (7): 559–70. <https://doi.org/10.1016/j.pt.2019.05.004>.
27. Boess, Franziska, Barbara Lenz, Juergen Funk, Urs Niederhauser, Simon Bassett, Jitao David Zhang, Thomas Singer, and Adrian B. Roth. 2017. "Use of Early Phenotypic in Vivo Markers to Assess Human Relevance of an Unusual Rodent Non-Genotoxic Carcinogen in Vitro." *Toxicology* 379 (March): 48–61. <https://doi.org/10.1016/j.tox.2017.01.018>.
28. Feng, Yan, Timothy J. Mitchison, Andreas Bender, Daniel W. Young, and John A. Tallarico. 2009. "Multi-Parameter Phenotypic Profiling: Using Cellular Effects to Characterize Small-Molecule Compounds." *Nature Reviews Drug Discovery* 8 (7): 567–78. <https://doi.org/10.1038/nrd2876>.
29. Jones, Lyn H., and Mark E. Bunnage. 2017. "Applications of Chemogenomic Library Screening in Drug Discovery." *Nature Reviews Drug Discovery* 16 (January): 285–296. <https://doi.org/10.1038/nrd.2016.244>.
30. Kulesa, Anthony, Jared Kehe, Juan E. Hurtado, Prianca Tawde, and Paul C. Blainey. 2018. "Combinatorial Drug Discovery in Nanoliter Droplets." *Proceedings of the National Academy of Sciences* 115 (26): 6685–90. <https://doi.org/10.1073/pnas.1802233115>.

# References (continued)

31. Vlachogiannis, Georgios, Somaieh Hedayat, Alexandra Vatsiou, Yann Jamin, Javier Fernández-Mateos, Khurum Khan, Andrea Lampis, et al. 2018. "Patient-Derived Organoids Model Treatment Response of Metastatic Gastrointestinal Cancers." *Science* 359 (6378): 920–26. <https://doi.org/10.1126/science.aao2774>.
32. Wawer, Mathias J., Kejie Li, Sigrun M. Gustafsdottir, Vebjorn Ljosa, Nicole E. Bodycombe, Melissa A. Marton, Katherine L. Sokolnicki, et al. 2014. "Toward Performance-Diverse Small-Molecule Libraries for Cell-Based Phenotypic Screening Using Multiplexed High-Dimensional Profiling." *Proceedings of the National Academy of Sciences* 111 (30): 10911–16. <https://doi.org/10.1073/pnas.1410933111>.
33. Zhang, Jitao David, Erich Küng, Franziska Boess, Ulrich Certa, and Martin Ebeling. 2015. "Pathway Reporter Genes Define Molecular Phenotypes of Human Cells." *BMC Genomics* 16: 342. <https://doi.org/10.1186/s12864-015-1532-2>.
34. Antolin, Albert A., Joseph E. Tym, Angeliki Komianou, Ian Collins, Paul Workman, and Bissan Al-Lazikani. 200017. "Objective, Quantitative, Data-Driven Assessment of Chemical Probes." *Cell Chemical Biology* 0 (0). <https://doi.org/10.1016/j.chembiol.2017.11.4>.
35. Finotello, Francesca, and Barbara Di Camillo. 2015. "Measuring Differential Gene Expression with RNA-Seq: Challenges and Strategies for Data Analysis." *Briefings in Functional Genomics* 14 (2): 130–42. <https://doi.org/10.1093/bfgp/elu035>.
36. Anders, Simon, and Wolfgang Huber. 2010. "Differential Expression Analysis for Sequence Count Data." *Genome Biology* 11 (10): R106. <https://doi.org/10.1186/gb-2010-11-10-r106>.
37. Robinson, Mark D., and Gordon K. Smyth. 2007. "Moderated Statistical Tests for Assessing Differences in Tag Abundance." *Bioinformatics* 23 (21): 2881–87. <https://doi.org/10.1093/bioinformatics/btm453>.
38. Jesse Lipp, *Why sequencing data is modeled as negative binomial*, <https://bioramble.wordpress.com/2016/01/30/why-sequencing-data-is-modeled-as-negative-binomial/>
39. Stanton, Benjamin Z., Emma J. Chory, and Gerald R. Crabtree. 2018. "Chemically Induced Proximity in Biology and Medicine." *Science* 359 (6380): [eaa05902](https://doi.org/10.1126/science.aao5902). <https://doi.org/10.1126/science.aao5902>.
40. Tom Gregory, <https://www.quora.com/Is-Gini-coefficient-outdated>

# References (continued)

41. [EBI online tutorial of Reactome](#), DOI: 10.6019/TOL.Reactome\_exbp-t.2017.00001.1
42. Frey, Brendan J., and Delbert Dueck. 2007. "Clustering by Passing Messages Between Data Points." *Science* 315 (5814): 972–76. <https://doi.org/10.1126/science.1136800>.
43. RNA-seq: <https://en.wikipedia.org/wiki/RNA-Seq>, in particular the [figure](#) by Thomas Shafee (CC BY 4.0)
44. [Affinity Propagation algorithm visualized](#)
45. Bray, Mark-Anthony, Shantanu Singh, Han Han, Chadwick T. Davis, Blake Borgeson, Cathy Hartland, Maria Kost-Alimova, Sigrun M. Gustafsdottir, Christopher C. Gibson, and Anne E. Carpenter. 2016. "Cell Painting, a High-Content Image-Based Assay for Morphological Profiling Using Multiplexed Fluorescent Dyes." *Nature Protocols* 11 (9): 1757–74. <https://doi.org/10.1038/nprot.2016.105>.
46. Nonejuie, Poochit, Michael Burkart, Kit Pogliano, and Joe Pogliano. 2013. "Bacterial Cytological Profiling Rapidly Identifies the Cellular Pathways Targeted by Antibacterial Molecules." *Proceedings of the National Academy of Sciences* 110 (40): 16169–74. <https://doi.org/10.1073/pnas.1311066110>.
47. Ljosa, Vebjorn, Peter D. Caie, Rob ter Horst, Katherine L. Sokolnicki, Emma L. Jenkins, Sandeep Daya, Mark E. Roberts, et al. 2013. "Comparison of Methods for Image-Based Profiling of Cellular Morphological Responses to Small-Molecule Treatment." *Journal of Biomolecular Screening* 18 (10): 1321–29. <https://doi.org/10.1177/1087057113503553>.
48. Young, Daniel W., Andreas Bender, Jonathan Hoyt, Elizabeth McWhinnie, Gung-Wei Chirn, Charles Y. Tao, John A. Tallarico, et al. 2008. "Integrating High-Content Screening and Ligand-Target Prediction to Identify Mechanism of Action." *Nature Chemical Biology* 4 (1): 59–68. <https://doi.org/10.1038/nchembio.2007.53>.
49. Cox, Michael J., Steffen Jaensch, Jelle Van de Waeter, Laure Cougnaud, Daan Seynaeve, Soulaiman Benalla, Seong Joo Koo, et al. 2020. "Tales of 1,008 Small Molecules: Phenomic Profiling through Live-Cell Imaging in a Panel of Reporter Cell Lines." *Scientific Reports* 10 (1): 13262. <https://doi.org/10.1038/s41598-020-69354-8>.
50. Stark, Rory, Marta Grzelak, and James Hadfield. 2019. "RNA Sequencing: The Teenage Years." *Nature Reviews Genetics*, July, 1–26. <https://doi.org/10.1038/s41576-019-0150-2>.